Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Effects of Type of Nonreinforcement and Number of Alternative Responses in Two Verbal Conditioning Situations: E. D. NEIMARK 209

Extinction by Omission of Food as Related to Partial and Secondary Reinforcement: S. H. HULSE, JR. AND W. C. STANLEY 221

Stimulus Generalization as a Function of the Serial Position of the Stimulus During Prior Training: M. E. SHAW AND F. A. KING 228

Role of Response Variables in Recognition and Identification of Complex Visual Forms: H. W. HAKE AND C. W. ERIKSEN 235

The Interaction of Several Variables in Latent Learning: O. DESIDERATO 244

Effect of Various Motivational Techniques Upon Learning and Performance Tasks: G. H. ZIMNY 251

The Reinforcement Difference Limen (RDL) Function for Shock Reduction: B. A. CAMPBELL 258

Psychological Factors in Muscle-Action Potentials: EMG Gradients: W. W. SURWILLO ... 263

American Psychological Association

Vol. 52 No. 4 October 1956

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1956 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1956 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 52, NO. 4 OCTOBER, 1956

EFFECTS OF TYPE OF NONREINFORCEMENT AND NUMBER OF ALTERNATIVE RESPONSES IN TWO VERBAL CONDITIONING SITUATIONS[1]

EDITH D. NEIMARK[2]

Indiana University

In his original experiment on the conditioning of "verbal expectations" Humphreys (**10**) provided the procedural paradigm for a good deal of subsequent research (**7, 8, 11**). The procedure may be summarized briefly as follows: (*a*) a series of discrete trials is given in which the start of each trial is signaled by the presentation of a brief stimulus; (*b*) upon the appearance of the stimulus *S* is to predict the occurrence of a random event by making one of two available responses, A_1 or A_2; (*c*) following *S*'s prediction and independently of it, *E* presents one of the to-be-predicted events, E_1 or E_2, according to a predetermined random sequence in which each event has a fixed probability of occurrence, π_1 and π_2. The dependent variable is the proportion of A_1 responses in successive trial blocks.

The major concern of most studies in this area has been the effect of proportion and arrangement of E_1 and E_2 and development of a quantitative formulation for this experimental situation. With the exception of the first group of studies there have been few attempts to investigate the variables affecting behavior in this situation or to extend the available quantitative theories of verbal conditioning to encompass the effects of these additional variables. The present study attempts to fill both these gaps. Specifically, the effects of two variables have been investigated: (*a*) the number of available response alternatives, two or three; and (*b*) event sequences in which exactly one of the E_i need not occur on each trial, i.e., some of the trials are blank in the sense that none of the E_i occur. In addition, the verbal conditioning model developed by Estes and Burke (**6, 7**) has been extended to cover the effects of these variables.

An additional purpose of the present study has been to compare learning under the traditional procedure with that under a procedure in which

[1] This article is based on a dissertation submitted to the faculty of the Department of Psychology of Indiana University in partial fulfillment of the requirements for the Ph.D. degree, September, 1953. The author wishes to express her appreciation to Dr. W. K. Estes for his unfailing suggestions, criticisms, advice, and patience.

[2] Now at AFP & TRC, Lackland A.F.B.

the occurrence of the E_1 is contingent upon *S*'s response. Such a procedure is achieved when *S* is instructed to produce one of the E_1 rather than to predict its appearance. The results of this procedure will be presented in Exp. II.

Experiment I

Method

Apparatus.—The *S* was seated facing a large black screen (36 × 30 in.) at the base of which three telegraph keys were mounted. Three milk-glass panel lamps, E_1, E_2, and E_3, were mounted upon the screen, one above each of the telegraph keys. A fourth lamp, the signal, was centered at *S*'s eye level.

The durations of the signal, the E_1, and the interval between signal and E_1 were automatically controlled by means of a mechanical timer. The timer and two sets of three knife switches, each wired in circuit with one of the E_1, were located on *E*'s side of the screen. The *E* side of the screen also contained three 6-w. bulbs each of which corresponded to one of the telegraph keys. Recording of responses and programing of E_1 were done manually by *E*.

The experiment was run in a light-tight sound-shielded room under conditions of low illumination.

Procedure.—The *S*s were 160 volunteers from Introductory Psychology classes at Indiana University. The *S* was seated facing the screen and given the following instructions: "This is going to be like a guessing game. First this top light comes on as a signal to guess and then one of these lower lights may follow it. You are to guess which of the lower lights will go on following the signal. Your object is to guess correctly as many times as you can.

"Now here is what you are to do. Be sure you are seated comfortably with the index finger of each hand resting on one of the outside keys (with one finger resting on each of the three keys). When this top light goes on you are to press the key above which you think the light will appear. Sometimes no light may appear—that is all right—you are to keep on guessing. You may not, however, guess that no light will appear. You must always press one and only one key when the top light goes on."

Following the instructions *S* was given a series of six practice trials in which the order of occurrence of the E_1 was RLLRRL for *S*s with two responses available, and LCRRLC for *S*s with three responses available. After questioning to ensure his understanding of the procedure, *S* was given a series of training trials in which the signal was presented for 2 sec. immediately followed by presentation of the appropriate E_1 for 2 sec. Following a 2-sec. intertrial interval the signal reappeared.

TABLE 1

Summary of Number of Training Trials and Relative Frequencies of Occurrence, π_1, of E_1, E_2, and E_3 for Each of the Noncontingent Groups

Group		No. of Trials	Relative Frequency		
			π_1	π_2	π_3
2-Key	A	66	1.00	0	
	B	100	.66	0	
	C	100	.66	.34	
	D	100	.66	.17	
3-Key	A	66	1.00	0	0
	B	100	.66	0	0
	C	100	.66	.17	.17
	D	100	.66	.085	.085

The training procedures for each of the groups differed with respect to the number of training trials, 66 or 100; the number of available responses, two or three keys; and the relative frequencies, π_1, with which each of the E_1 was presented. Details of the treatment of each of the eight groups are summarized in Table 1. Assignment to groups was random within restrictions to be outlined in the next section. For *S*s in the two-key groups the center key was covered by a metal shield and only the two outer keys were available. The position of each of the E_1 and its corresponding A_1 was counterbalanced within each group.

Immediately following the training series and with no break in the procedure, each *S* received a series of 50 extinction trials in which none of the E_1 ever appeared. The results of the extinction series are presented elsewhere (12).

Preparation of random sequences of E_1.—Since evidence (7, 11) indicates that the response sequence for a given *S* is influenced by the preceding E_1 sequence, it was decided to evaluate the effect of this dependence by using ten different E_1 sequences in each group. For Groups A in which only one light, E_1, was presented on all 66 training trials all 20 *S*s in each group obviously received the same sequence. For the remaining groups the sequences were prepared as follows: (*a*) a column of 100 two-digit numbers was selected from a table of random numbers. For $n \leq 65$ E_1 was entered for the corresponding trial. For Groups 2B and 3B the remaining trials were blank, i.e., no other *E* appeared; while for Group 2C E_2 was pre-

sented on the remaining trials. (*b*) The non-E_1 trials were divided by means of the table of random numbers into E_2 and E_3 trials in the case of Group 3C, or E_2 and blank trials in the case of Group 2D. (*c*) For Group 3D all E_2 trials of the corresponding 2D sequence were divided into E_2 and E_3 trials. Thus the sequence of E_1's was the same for the B, C, and D groups, training sequences for these groups differing only with respect to the treatment on non-E_1 trials (see Table 1). Ten separate random sequences were prepared in this fashion. Essentially, then, Exp. I consists of 10 replications of the 4×2 factorial design. The *S*s were assigned to groups randomly with the restriction that a replication must be completed before the next could be begun.

Results

Theoretical considerations.—In a recent paper by Estes and Straughan (**7**) a statistical learning theory (**4, 6**) has been successfully applied to a verbal conditioning situation of the Humphreys type by the use of several relatively simple assumptions. The signal marking the start of each trial is conceptualized as a population of stimulus elements which is sampled by *S* on each trial. The symbol θ is introduced to refer to the proportion of the population of elements which constitutes the effective sample on a given trial. Each time the signal is presented *S* must predict the occurrence of one of two random events, E_1 or E_2, by making the appropriate response, A_1 or A_2. On a fixed proportion, π_1, of the trial series E_1 is presented; on the remaining proportion, $1 - \pi_1$, E_2 is presented. It is assumed that on E_1 trials the stimulus elements sampled will become associated with response class A_1 and on E_2 trials the elements sampled will become associated with response class A_2. The probability of a response of class A_1 on any given trial is determined by the proportion of the population of stimulus elements associated with that response class. For a group of *S*s the mean probability (relative frequency) of a response of class A_1 on the nth trial, $\bar{p}_1(n)$, has been shown to be

$$\bar{p}_1(n) = \pi_1 - (\pi_1 - \bar{p}_1(0))(1 - \theta)^n \quad (1)$$

where $\bar{p}_1(0)$ refers to the group mean initial probability of A_1 at the start of training.

As n becomes large the value of the second term in Equation 1 becomes smaller and approaches 0, leaving π_1, the proportion of trials on which E_1 occurs, as the final probability of responses of class A_1. This prediction has been supported by the results of a number of experiments (**7, 8, 10**). Moreover, since the asymptotic value of $\bar{p}_1(n)$ is a function of π_1 alone, it follows that if exactly one E_i is presented on each trial and E_1 is presented on π_1 of the trial series, then Equation 1 should apply regardless of the number of response alternatives, A_i, and corresponding E_i available. In the present study, then, there should be no difference in final level of A_1 responding between two- and three-key groups trained under comparable E_1 treatments. A difference between two- and three-key groups will be reflected, however, by a steeper slope in the learning curve for the three-key groups since $\bar{p}_1(0)$ should become lower in value as more responses are available.

The extension of Equation 1 to multiple response groups has assumed—as is usually the case in verbal conditioning—that exactly one E_i is presented on each trial. In the case of Groups B and D of the present study this is not the case. Each of the E_i is presented on a proportion π_i of the trials but since $\Sigma_i \pi_i < 1$, the remaining $1 - \Sigma_i \pi_i$ trials are blank. To account for the effect of blank trials we assume that the stimulus elements sampled on a blank trial will become associated with the response class of which a member occurs on that trial. Given that no E_i is presented on the nth trial the probability of response A_1 on the $n + 1$ trial will be

$$p_1(n+1) = (1 - \theta)p_1(n) + \theta_{p_1}(n) = p_1(n). \quad (2)$$

It can then be shown that for any number, α, of available response classes and corresponding E_i each of which occurs with relative frequency π_i the group mean probability of response A_1

on the nth trial is

$$\hat{p}_1(n) = \frac{\pi_1}{\Sigma_i \pi_i} - \left[\frac{\pi_1}{\Sigma_i \pi_i} - \hat{p}_1(0)\right] \times (1 - \theta \sum_{i=1}^{\alpha} \pi_i)^n. \quad (3)$$

Whenever $\sum_{i=1}^{\alpha} \pi_i = 1$ Equation 3 reduces to Equation 1. However, for $\sum_{i=1}^{\alpha} \pi_i < 1$ the final probability of A_1 will exceed π_1 and the slope of the learning curve will be more gradual than in the traditional verbal conditioning situation.

In the following section we shall be concerned with the group mean proportion of A_1 responses on successive 10-trial blocks; and more specifically with the group mean curves of these proportions and the final level of responding at the end of training. An expression appropriate for the consideration of group data in 10-trial blocks may be obtained by a method outlined elsewhere (7). The group mean proportion of A_1 responses on the mth block of 10 trials turns out to be

$$\bar{P}_1(m) = \frac{\pi_1}{\Sigma_i \pi_i} - \left[\frac{\pi_1}{\Sigma_i \pi_i} - \bar{P}_1(1)\right] \times (1 - \theta \Sigma \pi_i)^{10(m-1)}. \quad (4)$$

Final level of A_1 responding.—As an estimate of asymptotic group mean proportion of A_1 responses, $\bar{P}_1(\infty)$, we have taken the group mean total of A_1 responses on the last 20 trials of the training series. The mean total and *SD* of A_1 responses for each of the eight groups are summarized in Table 2. In order to determine whether this measure represents a stable final level of responding, a t for paired measures was computed between A_1 totals for the first and last halves of the block of 20 trials. In all cases the obtained values of t fall short of significance at the .05 level (Table 2). It appears reasonable to assume that a stable level of responding has been reached.

TABLE 2

GROUP MEANS AND STANDARD DEVIATIONS OF FINAL LEVEL OF A_1 RESPONDING AND t TESTS OF STABILITY OF RESPONSE

(All t values based on 19 *df*; $N = 20$)

Group		Mean	*SD*	t_{diff} (Last 10-First 10)
2-Key	A	19.90	.300	0
	B	19.60	1.158	− .809
	C	12.40	2.245	− 1.182
	D	15.85	1.905	1.239
3-Key	A	19.70	.781	0
	B	19.70	.714	− 1.000
	C	13.10	2.700	.880
	D	15.50	3.668	− 1.703

Although the study was designed to enable over-all group comparisons of final response levels by means of analysis of variance, failure to fulfill the assumption of homogeneity of variance makes the use of this procedure questionable. In the case of Groups A and B the final levels of A_1 responding are so stable and so similar that further comparison was deemed to be gratuitous. In the case of Groups C and D an analysis of variance was run despite failure to fulfill the assumption of homogeneity of variance (Table 3). The obtained F values for differences due to the 10 random sequences of E_1, and for differences due to training treatment (C vs. D) are significant at better than the .001 level. The F values for differences due to number of response alternatives and for all interactions fall short of significance at the .05 level. These findings suggest that when only E_1 is presented during training (Groups A and B) the asymptote of A_1 responding approaches unity whether or not blank trials are included in the training sequence. In the case of groups receiving 100

TABLE 3

ANALYSIS OF VARIANCE OF FINAL LEVEL OF A_1 RESPONDING FOR THE C AND D NONCONTINGENT GROUPS

Source	*df*	*MS*	*F*
Sequence (S)	9	24.946	4.90***
No. of keys (K)	1	.613	
C vs. D group (G)	1	171.113	33.63***
S × K	9	6.168	
S × G	9	5.113	
K × G	1	5.515	
S × K × G	9	5.401	
Within sequence	40	5.088	
Total	79		

*** Significant at $P \leq .001$.

trials, approximately 66 of which are E_1 trials, (B, C, D) the asymptotic probability of A_1 approaches .66 only under the traditional procedure, Group C. The greater the proportion of blank trials in the training series the greater the amount by which the asymptotic response probability exceeds π_1.

The differences in final level of A_1 responding among two- and three-key groups given comparable training treatments were further tested by means of *t* tests for independent groups (Table 4). None of the obtained *t* values is significant at the .05 level. The prediction that the asymptotic level of A_1 responding should be independent of the number of response alternatives available appears to be amply supported.

As an additional test of the accuracy of theoretical predictions of asymptotic level of A_1 responding, theoretical and observed asymptotes were compared by means of *t* tests (with the variance of each group taken as an estimate of population variance). With the exception of Group 3B, all obtained *t* values fall short of significance at the .05 level (Table 5). In view of the apparently slight mean difference of .3 responses in Group 3B, a significant *t* is somewhat surprising. Possibly the slight variance of this group, .51, is an underestimate of the population variance. In general, however, the extended model for verbal conditioning appears to provide a satisfactory description of final level of A_1 responding for all groups.

Mean learning curves.—The mean proportions of A_1 responses in successive 10-trial blocks have been plotted in Fig. 1 where the small black dots refer to the two-key groups and the large white dots to the three-key groups. Theoretical curves through the observed points have

TABLE 4

VALUES OF *t* FOR DIFFERENCES BETWEEN GROUP MEAN TOTALS OF A_1 RESPONSES IN THE LAST 20 TRAINING TRIALS

Group	I2A	I2B	I2C	I2D	I3A	I3B	I3C	I3D
II2A	.650				.769			
II2B		5.915***				.183		
II2C			.483				.591	
II2D				.488				2.210*
I3A	1.047				.118			
I3B		.303				4.460***		
I3C			.745				1.614	
I3D				.369				1.944

Note.—I refers to Non-contingent groups; II to the Contingent groups.
* Significant at $P = .05$.
*** Significant at $P = .001$.

TABLE 5

PREDICTED AND OBSERVED PROPORTIONS OF A_1 RESPONSES FOR THE LAST 20 TRIALS FOR THE NONCONTINGENT GROUPS AND t TESTS OF ACCURACY OF PREDICTION

Group		Observed Proportion	Predicted Proportion	$t_{diff.}$ (Pred.-Obs.)
2-Key	A	.995	1.000	1.453
	B	.980	1.000	1.504
	C	.620	.660	1.553
	D	.793	.795	.011
3-Key	A	.985	1.000	1.676
	B	.955	1.000	2.424*
	C	.655	.660	.016
	D	.775	.795	.476

* Significant at $P \leq .05$; $df = 19$.

been derived from Equation 4. In obtaining each theoretical curve three parameters must be evaluated: π_1, $\hat{P}_1(1)$, and θ. The values of π_1 are fixed by the experimental procedure (Table 1); values for $\hat{P}_1(1)$ have been obtained from the observed group mean proportion of A_1 responses during the first 10 trials; θ has been evaluated for each group by means of a statistical device outlined in detail elsewhere (7).

Although the theoretical curves appear to describe the data adequately the question of goodness of fit is largely academic. Each point represents the mean of 20 Ss for 10 trials and undoubtedly reflects artifacts due to averaging. Secondly, each point represents the mean of 10 separate sequences of E_i, a variable whose effect upon responding has been demonstrated in this experiment as well as others (7, 11).

Comparisons among group curves are complicated by differences in both

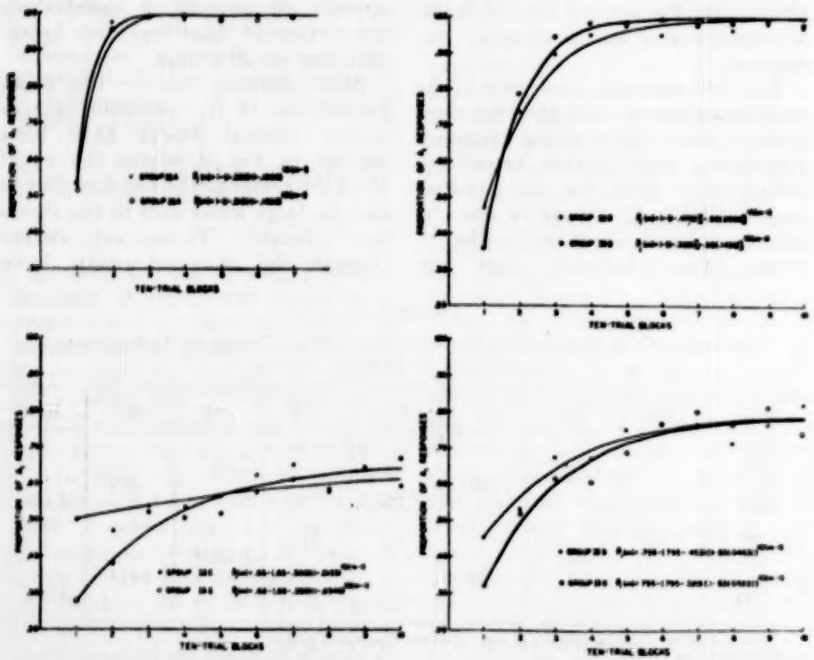

FIG. 1. Observed and theoretical mean proportions of A_1 responses in successive 10-trial blocks for the two- and three-key subgroups trained under each of the noncontingent treatments.

asymptote and initial value. However, in the case of Groups A and B, both of which have a unity asymptote, it is possible to test the prediction that the curves of the B groups should have a more gradual slope than those of the A groups (the prediction follows from the fact that the slope term for the A group contains the term $(1 - \theta)$ while the comparable term for the B groups is $(1 - .66\theta)$). Inspection of the curves for Groups A and B in Fig. 1 supports the prediction. Although the inclusion of blank trials in a training series of E_1 trials—a procedure somewhat analogous to partial reinforcement—does not affect the unity asymptote it does retard the rate of approach to that asymptote. The analogy to partial reinforcement is further strengthened by the finding of superior resistance to extinction for Group B (**12**).

Another assumption which follows from theoretical considerations is that the same value of θ should apply for all groups since the signal at the start of each trial is the same for all groups. An inspection of the θ values in Fig. 1 fails to support the prediction. Rather, the value of θ appears to be a decreasing function of the proportion of E_2 and E_3 trials. Similar results have been reported by Estes and Straughan (**7**) and our θ values are very similar to theirs for comparable groups. The only immediately apparent explanation for this finding has been suggested by Estes and Straughan: because of the massing of trials the stimulus elements present on successive trials cannot be assumed to constitute independent samples from the population of stimulus elements.

Experiment II

It has been noted above that in the traditional verbal conditioning situation S is required to predict the occurrence of random events which are presented with some fixed probability by E independently of S's response. A few Es (**2, 3, 9**) have deviated from this procedure by making the occurrence of the random event dependent upon the appropriateness of S's response; e.g., S is instructed to press the key which will "pay off" in poker chips, or turn on a light, etc. We shall refer to this procedure as the contingent case in order to distinguish it from the noncontingent or traditional procedure. Studies employing the contingent procedure with two available response classes, one of which is followed by E on approximately π_1 of the trial series while the other is followed by E on $1 - \pi_1$ of the trials, have obtained results very similar to those for the comparable noncontingent procedure (**2**). Assuming that trials on which E follows A_1 or no E follows A_2 function as E_1 trials, and that trials on which E follows A_2 or no E follows A_1 function as E_2 trials, then it is obvious that the contingent and noncontingent procedures should be equivalent. Whether the equivalence still obtains with the inclusion of blank trials or of more than two available response classes is open to question.

Starting from the assumptions outlined in Exp. I and making use of the additional assumption above that both E following A_1 and no E following A_2 are equivalent to an E_1 trial, it is possible to derive an expression for the probability of an A_1 response of the nth trial, $\bar{p}_1(n)$

$$\bar{p}_1(n) = \frac{1 - \pi_2}{2 - \pi_1 - \pi_2} - \left[\frac{1 - \pi_2}{2 - \pi_1 - \pi_2} - \bar{p}_1(0) \right] \times (1 - \theta(2 - \pi_1 - \pi_2))^n. \quad (5)$$

Essentially the same expression has been derived by Bush and Mosteller (1), as a special case of their linear operator model, and by Estes (5).

It will be noted that when $\pi_2 = 1 - \pi_1$ Equation 5 reduces to Equation 1, the expression for the noncontingent case. With the inclusion of blank trials, however, the equivalence will not generally be expected to hold.

Method

Apparatus.—The apparatus was the same as that in Exp. I; both experiments were run concurrently.

Subjects.—The *S*s were 160 volunteers from Introductory Psychology classes at Indiana University and were randomly assigned to one of eight groups (analogous to the groups of Exp. I) within the restrictions which applied in Exp. I.

Procedure.—The *S*s were run individually under a procedure identical with that of Exp. I except for necessary differences in instructions and in the presentation of E_1, E_2, E_3 or blank trials. Instructions for the contingent groups were similar to those of the noncontingent groups with the exception that *S* was told to press the key which he thought would light the lamp above it rather than to press the key below the light he thought would follow. After the recitation of the instructions, *S* received the series of six practice trials. In this instance *S* was told which key to press on each of the practice trials, but the series was otherwise identical to that used for the noncontingent groups. Following the correction of any mistake *S* made during practice, the training series was begun. The number of trials and treatments for each of the eight groups are summarized in Table 6. In this table π_1 refers to the proportion of trials on which the light above Key A_1 can be turned on if *S* presses that key; similarly for π_2 and π_3. Immediately upon the completion of the training series *S* received 50 extinction trials in which no key would light the lamp above it. The results of the extinction series are reported elsewhere (12).

Preparation of random sequences.—As was noted above, E_1 refers to trials on which the lamp above Key A_1 can be lighted if *S* presses Key A_1 and the relative frequency of such trials is denoted by π_1. Ten different sequences of E_1 were prepared for each group in a manner similar to that employed in Exp. I. For the A groups only one lamp could be lighted on each of the 66 training trials. In the case of the 2B and 3B groups the same sequences were employed as for the comparable noncontingent groups. For the C and D groups, the sequences of E_1 trials were identical to those of the comparable noncontingent groups but E_2 and E_3 trials were independently assigned; each with its appropriate probability, in the same manner as E_1 trials had been assigned. Thus, in Group 2C, e.g., over the entire series of trials the lamp above Key A_1 could be lighted about 66% of the time and the lamp above Key A_2 could be lighted about 34% of the trials; but on any given trial, either E_1, E_2, both or neither might be set to light.

TABLE 6

SUMMARY OF NUMBER OF TRAINING TRIALS AND RELATIVE FREQUENCIES OF OCCURRENCE, π_1, OF THE E APPROPRIATE TO EACH RESPONSE CLASS, A_1, FOR EACH OF THE CONTINGENT GROUPS

Group		No. of Trials	Relative Frequency		
			π_1	π_2	π_3
2-Key	A	66	1.00	0	
	B	100	.66	0	
	C	100	.66	.34	
	D	100	.66	.17	
3-Key	A	66	1.00	0	0
	B	100	.66	0	0
	C	100	.66	.34	.34
	D	100	.66	.17	.17

Results and Discussion

Final level of A_1 responding.—As in the case of the noncontingent groups, the group mean proportion of A_1 responses in the last 20 trials of the training series was taken as an estimate of the final probability of A_1 responding for each group. The mean totals and *SD*s of A_1 responses in the last 20 trials for each group are summarized in Table 7. The stability of the estimates has been tested for each group by means of a *t* for paired measures comparing A_1 totals in the first and second halves of the block of 20 trials. All resulting *t* values fall short of significance at the .05 level with the exception of that for Group 3D in which the A_1 total during the second half is significantly lower than during the first half of the 20-trial block.

TABLE 7

GROUP MEANS AND STANDARD DEVIATIONS OF FINAL LEVEL OF A_1 RESPONDING AND t TESTS OF STABILITY OF RESPONSE

(All t values based on 19 df; $N = 20$)

Group		Mean	SD	$t_{diff.}$ (Last 10-First 10)
2-Key	A	19.70	1.308	− 1.0
	B	16.45	2.012	− 1.926
	C	12.00	2.828	.212
	D	15.50	2.480	− .112
3-Key	A	19.80	.276	0
	B	16.60	2.939	0
	C	11.35	3.877	1.715
	D	13.00	4.242	− 2.269*

* Significant at $P \leq .05$.

Final probability of A_1 responding approaches unity only in the case of the continuously reinforced A groups. The differences among group mean A_1 totals of the B, C, and D groups were tested by means of analysis of variance (Table 8). Only the F for treatments reaches significance ($P < .001$) whereas F ratios for E_1 sequences, number of alternative responses, and all interactions fall short of significance at the .05 level. The differences between final A_1 levels for two- and three-key groups trained under comparable conditions were further tested by means of t tests (Table 4). All values of t fall short of significance at the .05 level except that for the D groups. However, since it is questionable whether Group 3D has reached a stable asymptote, the meaning of the significant t is not clear. In general, it would appear that the final level of A_1 responding is a decreasing function of the proportion of trials on which an alternative response is reinforced. Furthermore, final level of responding is independent of the number of response alternatives—at least at the level of two vs. three. These two findings are similar to those for the noncontingent procedure. Under the contingent procedure, however, responding seems to be less dependent upon E_1 sequence. This is not too surprising since the exact E_1 sequence for any given S is dependent upon his responses. Thus, 20 E_1 sequences rather than 10 have been compared and the resulting statistical test loses in sensitivity through inflation of the error term.

For additional comparisons of the two procedures the final levels of A_1 responding for contingent and noncontingent groups given comparable training treatments have been tested by t tests (Table 4). The nonsignificant t values for both two- and three-key A and C groups support Detambel's (2) finding that under the contingent analogue of the traditional procedure in which an E_1 occurs on each trial the final proportion of A_1 responses approaches π_1 just as it does under the noncontingent procedure. Moreover, it suggests that the equivalence of the contingent and noncontingent procedures, in the traditional type of situation, still holds when more than two responses are available. When blank trials are substituted for E_2 (or E_3) trials, the similarity of the contingent and non-

TABLE 8

ANALYSIS OF VARIANCE OF FINAL LEVEL OF A_1 RESPONDING FOR THE B, C, AND D CONTINGENT GROUPS

Source	df	MS	F
Sequences (S)	9	12.530	1.20
No of keys (K)	1	30.000	2.88
B vs C vs D group (G)	2	235.525	22.58***
S × K	9	5.500	
S × G	18	12.497	1.20
K × G	2	18.475	1.77
S × K × G	18	10.225	
Within sequences	60	10.433	
Total	119		

*** Significant at $P \leq .001$.

contingent procedures breaks down, as is indicated by the significant (.001) t values for all B groups. In the mixed series, Groups D, containing E_2 (E_3) and blank trials the nonsignificant t leaves the role of the blank trials in doubt. In general, however, the blank trial does not appear to play the same role in the contingent procedure that it does in the noncontingent.

Theoretical predictions of final level of A_1 responding for each of the two-key contingent groups have been computed from Equation 5 and are summarized in Table 9. Theoretical and observed final levels for each group were compared by means of t tests. In the case of the A and C groups the resulting t values fall short of significance at the .05 level; the t values for the B and D groups are significant at the .05 and .01 levels, respectively. It appears that the model for the contingent case is adequate for the traditional procedure in which one E_1 appears on each trial. For this procedure predictions of no difference between comparable contingent and noncontingent asymptotes are also supported by the data. However, with the inclusion of blank trials, the predicted asymptotes are significantly too low. This finding seems to challenge the assumption that the blank trials function as reinforcements of an alternative response. The rejection of this assumption, however, raises the question as to how S discriminates a blank trial from, e.g., a trial on which no light appears following an A_1 but light would have appeared following A_2. Since S is instructed that no light may appear on some trials it is possible that he develops the seemingly impossible discrimination, especially in the case of the B groups for whom A_2 (A_3) never produces its E. Conceivably the asymptotic A_1 probability for the B groups might approach unity after a very long training series, although there is no indication that it will do so in the data of this study. Clearly the development of an extended model for the contingent case will be more complex than was true of the noncontingent case. However, since the series containing blank trials in this experiment are more closely analogous to partial reinforcement in other types of experimental situations the problem merits further investigation.

Mean learning curves.—The mean proportions of A_1 responses in successive 10-trial blocks for the 2A, 2B and 2C contingent and noncontingent groups have been plotted in Fig. 2. Theoretical curves through the observed points of the noncontingent groups (solid curves and white points) have been computed from Equation 4; theoretical curves for the A and C contingent groups (broken lines and black points) have been computed from Equation 5. In comparing the contingent and noncontingent procedures for both A and C groups, not only the asymptotes but also the course of the learning curves appear to be almost identical. These findings further support theoretical predic-

TABLE 9

PREDICTED AND OBSERVED PROPORTIONS OF A_1 RESPONSES FOR THE LAST 20 TRIALS OF THE 2-KEY CONTINGENT GROUPS AND t TESTS OF ACCURACY OF PREDICTION

Group	Observed Proportion	Predicted Proportion	t_{diff} (Pred.-Obs.)
A	.985	1.000	1.000
B	.823	.746	2.324*
C	.600	.660	1.849
D	.775	.709	3.312**

* Significant at $P \leq .05$; $df = 19$.
** Significant at $P \leq .01$.

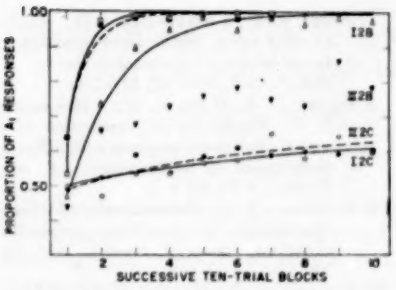

FIG. 2. Comparison of the learning curves of the two-key A, B, and C noncontingent groups (I) with curves of groups trained under the comparable contingent (II) procedure.

tions. An additional similarity among the contingent and noncontingent curves is reflected in the obtained values of the parameter θ: .1415 and .0183 for the A and C contingent groups, respectively, as compared with .1950 and .0135 for the noncontingent groups. Again the higher θ values for the A groups accords with previous findings.

Although no theoretical curve was plotted through the observed points of the contingent B group because of the failure of Equation 5 to describe the asymptote for this group, a comparison of the empirical curves suggests the possibility of a steeper slope for the contingent group than for the noncontingent group. This comparison is complicated, however, by a lower asymptote for the curve of the contingent group and by doubts as to the eventual final level for this group after a much longer series of trials. If the preceding analysis of behavior in this group is correct, a more or less sudden jump toward a unity asymptote would be expected in the contingent B curve with a prolonged training series.

Although a comparison of the curves of the three-key contingent and noncontingent groups has not been included, it is in general similar to that of the two-key groups in Fig. 2. With the exception of a lower initial level of A_1 responding the curves of the three-key contingent groups are very similar to those for the comparable two-key contingent groups. Similarities of the contingent three-key A and C groups' curves to their noncontingent counterparts are, however, somewhat obscured by chance differences in initial level of A_1 responding.

SUMMARY

Sixteen groups of 20 *Ss* each were run in an individualized modification of Humphreys' (10) verbal conditioning situation in which one response class, A_1, was most frequently "reinforced." Treatment of the groups differed with respect to: (*a*) number of response alternatives, two or three; (*b*) proportion of trials in which responses other than A_1 were "reinforced"; (*c*) whether "reinforcement" of a given response class was independent of or contingent upon *S*'s response. The purpose of the two experiments reported was to investigate the effects of the above variables upon relative frequency of A_1 responding and to test predictions from a mathematical model for a wider variety of verbal conditioning situations.

Results for the noncontingent groups indicate: (*a*) final level of A_1 responding is a decreasing function of the proportion of trials on which an alternative response is "reinforced" and is (*b*) independent of the number of alternative responses. (*c*) Inclusion of blank trials on which no response is "reinforced" depresses the slope of the learning curve. (*d*) Theoretical predictions concerning asymptotic levels of A_1 responding and learning rates were supported, but (*e*) predictions concerning parameters of the learning curves were not completely supported.

For the contingent groups: (*a*) final level of A_1 responding is a decreasing function of the proportion of trials on which an alternative response is reinforced and is (*b*) independent of the number of response alternatives. (*c*) Performance is equivalent to that under the noncontingent procedure only for groups in which some response is reinforced on each trial, while (*d*) the inclusion of blank trials results in lower final levels than under the comparable noncontingent procedure. (*e*) A mathematical model for the two-choice groups yielded accurate

predictions of asymptotes only for groups receiving no blank trials; (*f*) predictions concerning groups receiving blank trials were not supported.

References

1. Bush, R. R., & Mosteller, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.
2. Detambel, M. H. A reanalysis of Humphreys' "Acquisition and extinction of verbal expectations." Unpublished master's thesis, Indiana Univer., 1950.
3. Detambel, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, **49**, 97–105.
4. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
5. Estes, W. K. Individual behavior in uncertain situations. In C. L. Coombs, R. M. Thrall, and R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954.
6. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
7. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
8. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
9. Goodnow, J. J. Determinants of choice distributions in two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.
10. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
11. Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.
12. Neimark, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. Unpublished doctor's dissertation, Indiana Univer., 1953.

(Received September 12, 1955)

Journal of Experimental Psychology
Vol. 52, No. 4, 1956

EXTINCTION BY OMISSION OF FOOD AS RELATED TO PARTIAL AND SECONDARY REINFORCEMENT[1]

STEWART H. HULSE, JR. AND WALTER C. STANLEY

Brown University

Theoretical explanations of the fact that primary reinforcement[2] on every training trial usually generates less resistance to extinction than primary reinforcement on less than 100% of the training trials have recently involved two concepts. The first of these concepts is stimulus generalization decrement (**2, 4**). The argument is straightforward. Partial reinforcement during training cues the response to a stimulus pattern which involves the aftereffects of nonreinforcement as well as the aftereffects of reinforcement. In contrast, 100% reinforcement cues the response to a stimulus pattern which includes only the aftereffects of reinforcement. Generalization from stimuli present during training to stimuli present during extinction, when primary reinforcement is omitted altogether, should therefore be greater after partial reinforcement than after 100% reinforcement.

The major deduction from the generalization decrement hypothesis is this: The superior resistance to extinction generated by partial reinforcement should disappear if training trials are sufficiently distributed in time to permit dissipation of the perseverating aftereffects of nonreinforcement in nonarticulate *Ss* such as rats. Sheffield's (**4**) test of this deduction seemed to provide empirical support for its validity, but more recent investigations (**7, 9**) have clearly shown that the deduction is not generally valid.

The second concept which has been involved with the partial reinforcement effect is secondary reinforcement (**1, 3**). The argument here seems to be that the partial effect is dependent, in part at least, on the presence of secondary reinforcers during those training trials when primary reinforcement itself is omitted. However, this interpretation implies that secondary reinforcement produces a stronger habit than the specific primary reinforcement upon which that secondary reinforcement is based, a somewhat dubious assumption.

It is quite conceivable, nevertheless, that some factor *correlated* with the presence of secondary reinforcers during training trials involving a partial reinforcement regimen may account for the partial reinforcement effect. Since a check of the extant literature revealed that no study has systematically varied both ratio of reinforcement and the amount of secondary reinforcement, the present study was designed to begin fulfilling this need in the hope that a detailed analysis of the data might indicate that some general factor does underlie the partial reinforcement effect.

[1] About two-thirds of the data reported in the present paper constituted the first author's M.S. thesis; the remaining data were gathered as a continuation of the experiment which was aided by the award by Brown University of a Special Scholarship to the first author. The second author served as advisor.

[2] Throughout this paper the words reinforcement, secondary reinforcement, reinforcer, or secondary reinforcer, should be understood as devoid of any implication of the "mechanism" or genuine source of strengthening involved in learning. Such words will be used only in their functional or empirical sense.

METHOD

Subjects and Apparatus

The *Ss* were 78 experimentally naive, male albino rats, descendants of the Sprague-Dawley strain. The *Ss*' ages on the first day of training ranged from 62 to 76 days, the mean age being 67.4 days.

The inside dimensions of the enclosed alleyway were the same as those of Sheffield's (4) apparatus. The alleyway was 48 in. long and 4 in. wide. The start box was 6 in. wide and 7 in. long; the goal box was 8 in. wide and 12 in. long. Inside height throughout was 9 in. The start and goal boxes had hinged glass lids; the alleyway was covered by a hinged, screened frame.

A food compartment, 3 in. wide and 2½ in. deep, was attached to the far end of the goal box. The opening into the food compartment was 5 in. high, and the compartment's ceiling sloped downward to a height of 2½ in. at the rear so that *S* could not get entirely within the compartment. A metal food cup was screwed to the food compartment floor. A black curtain hung diagonally across the goal box.

The purpose of the separate food compartment was to provide a simple method for maximizing and minimizing secondary reinforcement. By opening a guillotine door which separated the goal box from the food compartment, *E* could expose *S* to cues most closely associated with *S*'s eating; by leaving the food compartment door closed *S*'s exposure to such cues could be reduced or minimized.

Guillotine doors were also located at the entrance to the alleyway and 4 in. from the end of the alleyway, and could be lowered to prevent retracing. The entire apparatus was painted flat black with the exception of the start box and the food cup which were painted white. The apparatus was lighted by four 15-w. bulbs suspended 2 ft. above the alleyway.

Hinged floors in the start box and goal box along with appropriate electrical circuits automatically controlled two Springfield timers to measure *S*'s start time and total running time as later defined. A 10-sec. interval timer consisting of a siliconetic relay which controlled a neon pilot light helped assure consistency in the duration of goal-box confinements during training and extinction. Another timing device signalled to *E* the end of the 15-sec. between-trials cage confinement.

The between-trials cage consisted of a box 9 in. long, 4½ in. wide, and 9 in. deep and was covered by a hinged, Plexiglas lid. The box was painted flat black throughout except for its Plexiglas lid. The floor of the cage was pivoted on an axle in the middle so that the floor's being tilted by *S*'s movements activated an impulse counter and provided a measure of *S*'s between-trials activity.

TABLE 1

EXPERIMENTAL DESIGN

Group Designation	Reinforcement Schedule*	Secondary Reinforcement**	
		Training	Extinction
100% O-O	100%	Opened	Opened
100% O-C	100%	Opened	Stays closed
50% O-O	50%	Opened	Opened
50% O-C	50%	Opened	Stays closed
50% C-O	50%	Stays closed	Opened
50% C-C	50%	Stays closed	Stays closed

* On Day 2 of training. All *Ss* received 100% reinforcement on Day 1.

** As varied by *E*'s manipulation of the food-compartment door. If the door was opened, secondary reinforcement was assumed to be maximized; if the door remained closed, secondary reinforcement was assumed to be reduced or minimized.

Experimental Design

The experimental design included two levels of three variables: (*a*) schedule of food reinforcement (50% or 100%); (*b*) secondary reinforcement during training (maximized or minimized); and (*c*) secondary reinforcement during extinction (maximized or minimized). It can be seen in Table 1 that the over-all design was an incomplete factorial design with two 100% groups missing. The reason for this was that the food compartment door had to be opened to provide food reinforcement. Thus the 100% reinforcement regimen precluded minimizing secondary reinforcement during training with our technique.

Procedure

Taming.—All *Ss* were given one week's taming. Each day, each *S* was allowed 30 min. to eat 7 gm., dry weight, of wet mash made from ground Purina Lab Chow, and was freely handled by *E* during the taming periods. The food deprivation schedule brought *Ss*' body weights down to about 85% of normal by the time training was started.

The *Ss* were placed in individual cages following the last day of taming, where they were housed throughout the following three days of experimentation. The *Ss* were tamed and housed in a different room from that in which the experimental apparatus was located.

Pretraining (Day 1).—The *Ss* were given 10 pretraining trials with 100% reinforcement. A 1-min. adaptation period in the between-trials cage preceded the first trial. Following this

adaptation period, *S* was placed in the start box facing the door; after an interval of 2 to 3 sec., the start-box door was raised. The goal-box door was closed as soon as *S* depressed the floor section whose movement stopped the total running time clock. During Trials 1–4, the goal-box curtain was hung entirely to one side of the goal box; and the food-compartment door remained continuously open. On Trial 5, the curtain was hung so that it extended from one side wall of the goal box to 2 in. from the other side wall of the goal box. The curtain then remained in this position throughout training and extinction. On Trial 6 and thereafter, the food compartment door was kept closed and was opened on Day 1 only after *S* had passed the curtain and seemed oriented toward the food-compartment door. There was often a delay of some seconds before *S* passed the curtain, and consequently, before the food-compartment door was raised. Once *S* had learned to run into the goal box and to eat the bit of food quickly, its goal-box confinement became 10 sec. in duration. Confinement in the between-trials cage between successive trials was 15 sec. in duration.

Food reinforcement consisted of a wet mash pellet weighing about .08 gm., dry weight. Supplementary feeding came 1 to 2 hr. after *S* was run, and 22 to 24 hr. before *S*'s next day's training. The supplementary feeding brought each *S*'s daily ration to 7 gm.

A discard criterion to eliminate especially slow learners was set in advance. The criterion was *S*'s failure to eat the food within 10 min. from the start of any one trial during either pretraining or training. Thirteen *S*s met the discard criterion. These and six *S*s discarded either because of apparatus failure, extreme resistance to *E*'s handling, or difficulties in maintaining the food deprivation schedule, were replaced by assigning extra *S*s from a pool tamed for such replacement purposes.

Training (Day 2).—Thirty training trials were given with schedule of reinforcement and amount of secondary reinforcement varying according to the experimental design (Table 1). The 100% groups were reinforced with food on all training trials. The 50% groups were randomly reinforced with food on half the training trials according to the randomization procedure described by Sheffield (**4**). In respects other than schedule of reinforcement and secondary reinforcement, the training procedure on Day 2 was like that during the last five trials of pretraining.

Extinction (Day 3).—During the 30 extinction trials, secondary reinforcement was maximized by *E*'s opening the food-compartment door on all trials, or minimized by *E*'s leaving the food-compartment door closed on all trials. The interior of the apparatus was thoroughly washed prior to running *S*s. Food was omitted on all extinction trials. If *S* failed to reach the goal box in 2 min., it was removed and placed in the between-trials cage to await its next trial. Thus all *S*s received 30 placements in the start box regardless of their performance.

Measures of Performance

During both training and extinction *E* recorded to the nearest .1 sec. *S*'s start time (the time *S* took to leave the start box after *E* had raised the start-box door) and *S*'s total running time (the time *S* took to reach the goal box after *E* had raised the start-box door). Because the 2-min. nonresponse criterion during extinction rendered both of these measures indeterminate on some trials, individual performance measures which reflect start time and total running times during extinction had to be different from comparable measures which reflect performance during training. The differences in individual measures can be explained most communicatively and economically in the Results section.

A third measure of *S*'s behavior reflected *S*'s movements from one end to the other of the between-trials cage and was read from an impulse counter. This measure was not taken during the running of the first two replications of *S*s (12 *S*s).

Results

Training

Median start times and median total running times were analyzed for Day 1 alone, for Trials 1–30 on Day 2, and for Trials 16–30 on Day 2. Table 2 presents the medians of each of these measures with *S*s grouped according to the three conditions of training which were used on Day 2. The marked comparability in performance is apparent from inspection of the table.

Wilcoxon's (**8**) nonparametric test for unpaired replicates showed that no training group when compared with either of the other two groups differed sufficiently so that the *P* value was close to .10, both tails, during either Day 1 or Day 2. Further and more complex statistical analyses are therefore unnecessary to conclude that these findings provide no evidence that the three training

TABLE 2

MEDIANS OF MEDIAN RESPONSE TIMES FOR THE THREE TRAINING GROUPS DESIGNATED ACCORDING TO DAY 2 TRAINING CONDITIONS

Measure	Training Condition	Day 1 (Trials 1–10)	Day 2 (Trials 1–30)	Day 2 (Trials 16–30)
Start time (sec.)	100%	6.6	0.6	0.4
	50% O	8.1	0.6	0.4
	50% C	9.3	0.6	0.4
Total running time (sec.)	100%	22.2	3.6	2.0
	50% O	26.0	3.0	1.8
	50% C	20.0	2.9	1.8

groups and training conditions differed except by chance.

Extinction

Both start time and total running time were also analyzed to evaluate performance during extinction. Here a frequency measure was used. The measure was obtained by determining a grand median for the particular trials and response time being considered for all 78 *S*s combined. Next, for each *S* the number of trials faster than or equal to this grand median was determined. Such frequency scores were obtained for all 30 extinction trials and for only the last 15 extinction trials.

Table 3 gives the mean performance of the six extinction groups in terms of these frequency measures. Table 4 summarizes the analyses of variance which considered the two 100% groups and the two 50% groups for which secondary reinforcement was maximized during training as a 2 × 2 analytic unit. The results of these and subsequent analyses of variance are in terms of *t* rather than *F*. The reason for this is that for assessing the reliability of both the partial effect

TABLE 3

RESISTANCE TO EXTINCTION IN TERMS OF MEAN NUMBER OF RESPONSE TIMES FASTER THAN OVER-ALL MEDIAN RESPONSE TIME

Group	Start Time		Total Running Time	
	Trials 1–30	Trials 16–30	Trials 1–30	Trials 16–30
100% O-O	15.7	8.8	15.5	8.4
100% O-C	9.9	6.0	11.0	5.8
50% O-O	17.2	10.4	19.1	9.8
50% O-C	14.7	6.9	13.7	6.4
50% C-O	13.6	8.3	17.7	9.0
50% C-C	11.0	5.9	12.5	5.8
Over-all median	1.2	3.2	10.8	16.9

and the effect of secondary reinforcement on extinction performance, we are testing hypotheses of directionality. For such tests the one-tailed *P* value is appropriate, and this one-tailed *P* value is most simply obtained from a table of *t*.

Table 4 shows that when start times and total running times for Trials 1–30 are considered, the 50%

TABLE 4

SUMMARY OF STATISTICAL ANALYSES OF RESISTANCE TO EXTINCTION TAKING GROUPS 100% O-O, 100% O-C, 50% O-O, AND 50% O-C AS AN ANALYTIC UNIT

Source of Variation	Trials Analyzed	Values of *t*: Start Time	Values of *t*: Total Running Time
Schedule of reinforcement	1–30	1.83*	1.67*
	16–30	1.20	0.89
Secondary reinforcement during extinction	1–30	2.43**	2.65**
	16–30	3.08**	2.66**
Interaction	1–30	0.94	0.22
	16–30	0.30	0.41

* $.05 \geqslant P > .01$.
** $P < .01$.

TABLE 5

SUMMARY OF STATISTICAL ANALYSES OF RESISTANCE TO EXTINCTION TAKING GROUPS 50% O-O, 50% O-C, 50% C-O AND 50% C-C AS AN ANALYTIC UNIT

Source of Variation	Trials Analyzed	Values of t	
		Start Time	Total Running Time
Secondary reinforcement during training	1–30	1.96*	0.80
	16–30	1.47	0.66
Secondary reinforcement during extinction	1–30	1.38	3.24**
	16–30	2.79**	3.18**
Interaction	1–30	0.01	0.01
	16–30	0.51	0.01

* $.05 > P > .01$.
** $P < .01$.

schedule of reinforcement produced reliably greater resistance to extinction than the 100% schedule of reinforcement ($.05 \geqslant P > .01$). The table also shows that groups with maximized secondary reinforcement during extinction performed reliably better than the groups with minimized secondary reinforcement during extinction ($P < .01$). However, no significant interaction obtained between reinforcement schedule and secondary reinforcement conditions during extinction.

As a second analytic unit, we considered the four 50% groups. Table 5 summarizes the analyses of variance on these four groups. The table shows that minimizing secondary reinforcement during extinction reliably reduced resistance to extinction ($P < .01$), except with start time as the measure over all extinction trials. However, minimizing secondary reinforcement during *training* reliably reduced resistance to extinction ($.05 \geqslant P > .01$) only when start time is taken over all 30 extinction trials. The table also shows that no reliable interaction obtained between secondary reinforcement during training and secondary reinforcement during extinction.

The final unit of analysis (Table 6) considered the 100% groups and the 50% groups for which secondary reinforcement was minimized during training. Here only secondary reinforcement during extinction proved statistically reliable; the schedule of reinforcement and the interaction did not.

Between-Trials Activity

Between-trials activity was analyzed separately for Day 2 (training) and for Day 3 (extinction). The individual score for each *S* on each day was that *S*'s total impulse count for the day. It was found by t tests that only chance differences obtained between any two of the three conditions of training, and, by analysis of variance, that only chance differences obtained among the six extinction groups. It should be noted that

TABLE 6

SUMMARY OF STATISTICAL ANALYSES OF RESISTANCE TO EXTINCTION TAKING GROUPS 100% O-O, 100% O-C, 50% C-O, AND 50% C-C AS AN ANALYTIC UNIT

Source of Variation	Trials Analyzed	Values of t	
		Start Time	Total Running Time
Schedule of reinforcement	1–30	0.24	0.91
	16–30	0.26	0.25
Secondary reinforcement during extinction	1–30	2.08*	2.48**
	16–30	2.22*	2.28*
Interaction	1–30	0.78	0.17
	16–30	0.20	0.25

* $.05 > P > .01$.
** $P < .01$.

the mean impulse count for the 66 *Ss* for which activity data were available was 101 for Day 2 (training) and 95 for Day 3 (extinction).

Discussion

The first noteworthy aspect of our findings is the consistency with which the opening of the food-compartment door during extinction showed the functional properties of a secondary reinforcer. On the assumption that this would not have held if opening the food-compartment door had not been paired with *S*'s eating food, the data operationally support our label, secondary reinforcement, for the opening of the food-compartment door.

The second important aspect of our findings is that a reliable partial reinforcement effect was demonstrated when secondary reinforcement was present on every trial during training, but not when secondary reinforcement was omitted during the non-food-reinforced training trials. Although Rubin (3) has argued theoretically that this finding should hold, his experiment provided no evidence that the partial effect could be demonstrated even if secondary reinforcement during training had been maximized with his specific procedure and apparatus.

A third aspect of our findings provides further evidence that the presence of secondary reinforcers on every training trial helps assure the demonstration of a partial reinforcement effect. Holding schedule of food reinforcement constant at 50%, it was found that resistance to extinction was greater if secondary reinforcement was present on every trial than if secondary reinforcement was omitted on non-food-reinforced trials. It should be noted, however, that this difference is reliable with start time as the measure of performance during extinction, but not with total running time as the measure. It seems reasonable to suggest that start times probably reflected a more consistent and discrete class of responses from trial to trial then total running time. The start times, for example, reflected a very short temporal chain of locomotory movements while total running times reflected a great variety of locomotory and other movements, as starting to move, stopping, standing, retracing, etc., which consumed time before *S* actually reached the goal box during many extinction trials.

On the basis of the present findings, we suggest that any theory which would account for the partial effect must also explain the demonstrated role of secondary reinforcers. Specifically, we propose that during partially reinforced training *S* learns to eat food in the goal box when food is present, but learns to do something else when food is absent (e.g., *E* observed such responses as jumping and biting the curtain). Partially reinforced *Ss* thus enter extinction proper with something to do in the empty goal box, while continuously reinforced *Ss* have learned only to eat there. This, in turn, means that during extinction the partially reinforced *S*, by making these other responses, is quickly removed from cues stimulating eating. Thus, conditioned eating is protected from rapid extinction. In contrast, the continuously reinforced *S*, with no other responses immediately available, remains exposed to cues stimulating eating for a longer time. Thus conditioned eating should extinguish more quickly. But according to the frustration-generated excitement theory (5) of food reinforcement, this faster extinction of eating after continuous reinforcement should lead to faster loss of the instrumental response, and the partial effect should occur. Further, if strong as compared to weak secondary reinforcers (i.e., cues associated with eating) are present on non-food-reinforced training trials, greater excitement should be generated. This, in turn, should increase the probability that an alternative response will occur and be learned, and thus should increase the probability of demonstrating the partial effect.

Insofar as the between-trial activity can be regarded as reflecting the after-effects of reinforcement and nonreinforcement, the no-difference results serve

to question the generalization decrement hypothesis (**2, 4**). However, it should be noted that the small size of the between-trials cage probably rendered it difficult for *S* to display much activity during its 15-sec. confinement. That the technique was not very sensitive is further indicated by the fact that the mean impulse count per between-trials confinement was about 3. This meant that *S*s, on the arithmetic average, went back and forth only twice in the between trials cage, since the first counter impulse was usually due to *S*'s being placed in the cage by *E*.

Finally, we should mention that our three conditions of training produced no reliable differences in performance during training. This finding is consistent with a comparable finding by Rubin (**3**) and not consistent with some of the findings of Denny (**1**). Our best guess here is that time measures of performance in a straight alley probably reflect correlated motivational factors more than correct turn measures of performance in a T maze (cf., **6**).

Summary

The present experiment compared resistance to extinction by omission of food of a running response after partial and after continuous reinforcement when secondary reinforcement was varied both during training and during extinction. Thirteen rats were used in each of six extinction groups. The results were: (*a*) the three training conditions produced no reliable differences in performance during training; (*b*) opening a food-compartment door during extinction produced reliably greater resistance to extinction than leaving the door closed, thus justifying our label, secondary reinforcement, for the opening of the door; (*c*) partial reinforcement produced reliably greater resistance to extinction than continuous reinforcement only when secondary reinforcement was present on every training trial; (*d*) partial reinforcement with secondary reinforcement maximized during training produced reliably greater resistance to extinction than partial reinforcement with secondary reinforcement minimized during training, but only with start time and not with total running time as the measure of performance; (*e*) none of the possible interactions between variables proved reliable; and (*f*) no reliable differences in activity between trials were found, but the technique used to determine between-trials activity did not appear very sensitive. Following a discussion of the findings, it was suggested that a generalized explanation of the effect of partial reinforcement on resistance to extinction might be based on a frustration-excitement hypothesis concerning the "mechanism" of food reinforcement.

References

1. Denny, M. R. The role of secondary reinforcement in a partial reinforcement learning situation. *J. exp. Psychol.*, 1946, **36,** 373–389.
2. Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47,** 193–234.
3. Rubin, L. S. A demonstration of superior resistance to extinction following continuous reinforcement as compared with partial reinforcement. *J. comp. physiol. Psychol.*, 1953, **46,** 28–32.
4. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39,** 511–526.
5. Sheffield, F. D., & Campbell, B. A. The role of experience in the "spontaneous" activity of hungry rats. *J. comp. physiol. Psychol.*, 1954, **47,** 97–100.
6. Stanley, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43,** 249–260.
7. Weinstock, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47,** 318–322.
8. Wilcoxon, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1949.
9. Wilson, W., Weiss, E. J., & Amsel, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50,** 51–60.

(Received September 13, 1955)

Journal of Experimental Psychology
Vol. 52, No. 4, 1956

STIMULUS GENERALIZATION AS A FUNCTION OF THE SERIAL POSITION OF THE STIMULUS DURING PRIOR TRAINING[1]

MARVIN E. SHAW AND F. A. KING

The Johns Hopkins University

The general hypothesis with which this report is concerned is that erroneous responses which occur in the usual stimulus-response situation are determined in part by the direction and number of the associations which have been formed among the stimuli prior to training. Razran (**10, 11**) has shown that when forward associations presumably exist between two stimuli (e.g., the word complex "Yankee-Doodle"), significantly more generalization is observed when *S* is conditioned to the second stimulus and tested by presenting the first than when he is conditioned to the first and tested by presenting the second. He proposes a "subsequent" testing hypothesis to account for this finding. According to this hypothesis, stimulus generalization is generated at the time the test stimulus is presented. The following diagram shows our interpretation of what happens when a test stimulus is presented (if generalization occurs):

$$TS \rightarrow (R) = CS \rightarrow GCR$$

That is, the test stimulus (*TS*) evokes the response (*R*) which is the conditioned response and which serves as a stimulus for the generalized conditioned response (*GCR*).

Data congruent with this interpretation are also found in sensory preconditioning experiments. Wickens and Briggs (**15**) proposed that sensory preconditioning is a special case of mediated generalization and reported results interpreted as supporting this view. More recently, Silver and Meyer (**14**) have demonstrated that transfer in preconditioning experiments varies in amount as a function of the temporal relations between preconditioning stimuli. They found that forward presentation was superior to either backward or simultaneous presentation.

It is believed that a similar process is operative when an individual is trained to make a particular voluntary response to one stimulus and later commits the error of making this response when other stimuli are presented. The nature of the process intervening between the test stimulus and the occurrence of the erroneous response is still obscure. However, there is some reason to believe that this is an automatic process; i.e., that it does not depend upon *S*'s making a conscious response or judgment of any kind. In fact, if a conscious response or discrimination were made by *S*, it is probable that no erroneous responses would be made when the response can be controlled voluntarily. Some indirect experimental evidence for this view is available. Bugelski and Scharlock (**3**) and Russell and Storms (**13**) found that learning situations which provided for the possibility of mediated association resulted in significantly better learning than situations which did not provide for this possibility, even though *S*s could not verbalize the mediating terms nor were they able to report any ideational or perceptual use of the mediating material.

If this formulation is correct, then it should be possible to predict the course

[1] This research was done under Contract N5ori-166, Task Order I, between the Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-195, Project Designation No. NR 145-089, under that contract.

of stimulus generalization (erroneous responses) when the associations have been established among the stimuli by means of training procedures. One convenient training procedure for establishing forward associations among stimuli is the serial learning situation.[2] There seems to be reasonably good evidence from the results of serial order learning experiments (**1, 2, 5, 6, 9**) that during learning both direct and remote forward associations are formed among the stimuli which comprise the serial list. There is less acceptable evidence that associations are formed in the backward direction (**4, 9**). For purposes of the present investigation, this question is of no great importance so long as it is assumed that forward associations are stronger than corresponding backward ones, if such are formed. Therefore, the first item in the list should be the final member of *no* associated pair (except for a possible association with the last item in the list), the second item in the list should be the final member of *one* associated pair, the third item should be the final member of *two* associated pairs, and so on.[3]

On the basis of the above analysis, one would expect that if *S* is trained to make a certain response to one item from a previously learned serial list and other items from this list are later presented as stimuli, the first item in the list should be evoked less frequently than the second item in the list under comparable conditions (depending as it does upon possible backward associations or forward associations between it and items near the end of the list), the second item less frequently than the third, and so on. According to the mediated generalization hypothesis, then, when items of a learned list are subsequently used as stimuli in a conditioning (training) situation, the amount of stimulus generalization (erroneous responding) should be a function of the serial position of the item used as the training stimulus. This was the primary hypothesis which this experiment was designed to test.

In addition, responses to test stimuli could be examined with respect to their distance forward or backward from the training stimulus. If this is done, one might expect a gradient of generalization on either side of the training stimulus, but the crucial thing for our hypothesis is that the amount of generalization to test stimuli which had preceded the training stimulus in the original serial list should be greater than that to test stimuli which had followed the training stimulus in the original serial list.

Method

Apparatus.—Two pieces of equipment were used in this experiment. One was a standard memory drum which was used during the initial training period to establish the serial associations among the nonsense syllables which were used later for the conditioned response.

The other unit was used for the conditioning situation. A stylus was mounted on a table surface in such a way that it could be moved in the horizontal plane. The stylus was connected to a pencil which rested upon a recording surface. When the stylus was moved, the pencil attachment recorded the amount of movement upon the recording paper, thus giving a measure of frequency and amplitude of response. To the left and just above the level of the table there was a ground-glass screen upon which the stimuli (nonsense syllables) could be projected. When a syllable was projected on the screen, a timer was started automatically. Any slight movement of the stylus from the starting position would stop the timer, thus giving a measure of response latency. Duration of the stimulus presentation was controlled by an electronic timing device.

Subjects.—Seventy-two high-school girls, ranging in age from 17 to 19 yr., served as *Ss*. They were recruited by *E* and were paid for their services. All *Ss* were right-handed.

[2] Another training procedure that could be used is the paired-associates method. The serial learning situation was selected because it permits the formation of a varying number of forward associations among pairs of items in the serial list.

[3] This analysis may be expected to break down toward the end of the list since it is not known just how much two syllables can be separated and still become associated through learning (although Bugelski [**2**] reports at least six). Thus it might be expected that at some point in the list the number of associations would become constant for all syllables which lie beyond that point.

Procedure.—During the first part of the experiment (hereafter referred to as original learning), each *S* was required to learn a list of six nonsense syllables (TOV, GAH, YIL, ZUK, XEJ, and QIF) by the method of anticipation. The syllables were presented by means of a standard memory drum. Each syllable was in view for 20 sec., and the intersyllable interval was 20 sec. Each trial was separated from the next trial by an interval of 60 sec., and *E* made it clear to *S* that the trials were separate and distinct. Learning was continued to a criterion of eight successive correct trials. The procedure was exactly the same for all *S*s except that the order of presentation was varied so that each syllable appeared in each serial position for an equal number of *S*s. This procedure resulted in six groups, each of which learned the syllables in a different order.

Following the completion of original learning, *S* was given a rest period of approximately 3 min. Then she was seated before the conditioning apparatus so that the stylus could be grasped easily with the right hand and so that the ground-glass screen was at eye level. The *S* was told that a certain syllable would be flashed upon the screen (the designated syllable was shown for a few seconds) and that she was to move the stylus to the left as rapidly as possible whenever this syllable appeared. Speed was emphasized; *S* was told that if she did not respond rapidly enough, *E* would sound a buzzer. No inhibitory instructions were given, i.e., no mention was made of what *S* was to do if any syllable other than the designated one should appear on the screen.

When *E* was sure that *S* understood the instructions, *S* was given 50 training trials to the training syllable. For one group the training stimulus had been in the first position during original learning; for a second group the training stimulus had been in the second position during original learning, etc. Thus there were six groups of 12 *S*s each; each group was trained to a syllable which had occupied a different position in the serial list during original learning. Two *S*s from each of the original learning groups were included in each test group in order to counterbalance stimulus syllables. Duration of the stimulus was .85 sec., and the interstimulus interval varied from 5 to 20 sec. Buzzer reinforcements were given at random. At the end of 50 training trials, and without any warning to *S*, a series of test trials to other syllables was given. The stimuli for this series consisted of 20 test syllables (four presentations of each of the other five syllables which were used in the original learning situation) interspersed among 35 additional presentations of the training stimulus.

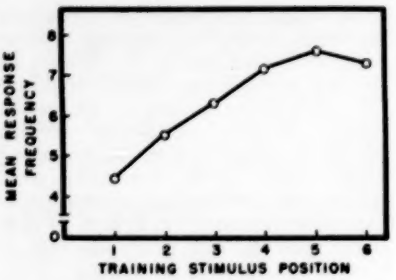

FIG. 1. Mean frequency of response to test stimuli as a function of the serial position of the training stimulus during original learning.

RESULTS

Before presenting the results related to the hypothesis being tested, it is necessary to inquire whether the six test groups were equivalent in learning ability as measured by their performance during original learning. The over-all mean of number of trials to criterion was 14.72, $SD = 4.00$. The means for Groups 1 through 6 were, respectively, 14.75, 14.42, 15.92, 13.75, 15.50, and 14.00. Simple analysis of variance yielded an F of 0.80, which, with 5 and 66 *df*, does not approach significance ($P > .20$). Also, the test groups appear to be equivalent with respect to response latency during the 50 training trials. Means of response latency were .33, .34, .33, .35, .34, and .36 for Groups 1 through 6, respectively. These means did not differ significantly ($F = 1.07$, $P > .20$).[4]

Frequency.—Frequency scores were obtained by counting all of the responses made to the test syllables by each *S*. In Fig. 1 mean frequency of response is shown as a function of the serial position of the training stimulus

[4] During the training period all *S*s responded 100% of the time to the training stimulus with a maximum response; hence, the question of equivalence of test groups with respect to frequency and magnitude scores does not arise.

during original learning. Each plotted point represents the mean number of erroneous responses made by the 12 *Ss* in each group. It can be seen that in general frequency of response increased as the syllable used as the training stimulus was chosen from successive positions within the serial list, in agreement with our hypothesis. Over-all differences among groups were statistically reliable ($\chi^2 = 13.48$; $df = 5$; $P < .02$). However, the trend is of greater interest for the hypothesis under consideration. A simple test for linearity of regression of frequency scores on serial position was statistically significant ($P < .01$).

Frequency scores were also analyzed as a function of their distance forward and backward from the training stimulus in the original serial list. The results of this analysis are shown in Fig. 2. This shows a relatively flat gradient of generalization on either side of the training stimulus, with the portion resulting from responses to test stimuli which had preceded the training stimulus during original learning being higher than the portion resulting from responses to test stimuli which had followed the training stimulus during original learning. Unfortunately the various points plotted in Fig. 2 are represented by different and overlapping subgroups (e.g., the two points adjacent to the training stimulus are for 60 *Ss* each, the next nearest two for 48 *Ss* each, etc.); hence, there appears to be no adequate way of statistically evaluating these data. However, 48 *Ss* were tested both to stimuli which had immediately preceded and to stimuli which had immediately followed the training stimulus during original learning. These *Ss* responded 38% of the time to the test stimuli which had immediately preceded the training stimulus as compared with 27.6% of the time to test stimuli which had immediately followed the training stimulus. This difference is statistically reliable ($\chi^2 = 4.72$; $P < .05$).

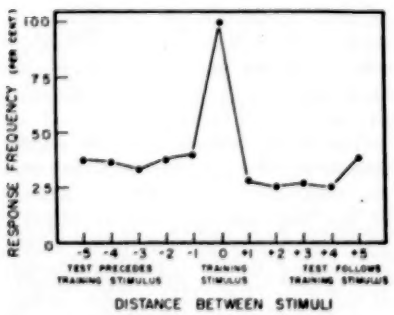

FIG. 2. Mean response frequency as a function of the distance of the test stimuli (forward and backward) from the training stimulus in the original learning list.

It is also interesting to note the relative increase in frequency of response to test stimuli which were at the extremes with respect to distance from the training stimulus. This could have been a chance occurrence, but it seems more likely that this is due to forward associations which were formed between the syllables near the end of the list and syllables near the beginning of the list, despite *E*'s efforts to keep trials separate and distinct.

Taken together, the frequency data agree reasonably well with the hypothesis that generalized responses are a function of the direction and number of associations which have been formed among the stimuli prior to training.

Magnitude.—Amount of generalization was also measured in terms of magnitude, where magnitude is defined as the total amount of movement divided by the total number of times the test stimuli were presented. The

maximum amount of movement possible was 2.7 in. Since *S*s always gave a maximum response to the training stimulus, the mean magnitude of response to the training stimulus was 2.7 with $SD = 0$.

Figure 3 shows the mean magnitude of erroneous response as a function of the serial position of the training stimulus during original learning. The differences among group means are small and are statistically unreliable, but again interest centers in the trend of the function. Linearity of regression was statistically significant ($P < .05$). In agreement with the frequency data, this result supports the hypothesis.

Figure 4 shows the mean magnitude of response to test stimuli as a function of their distance forward and backward from the training stimulus in the original serial list. The form of the plot is essentially the same as the corresponding plot of frequency data, and the same statistical difficulties exist. Again considering the response to test stimuli which had immediately preceded and to test stimuli which had immediately followed the training stimulus during the original learning, we find significantly ($t = 2.14$; $P < .05$) more generalization to the former (mean = 5.94 in.) than to the latter (mean = 4.64 in.). These data are in agreement with the hypothesis under consideration.

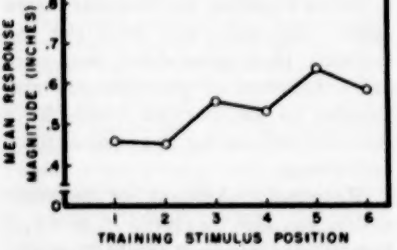

FIG. 3. Mean magnitude of response (inches of movement) to the test stimuli as a function of the serial position of the training stimulus during original learning.

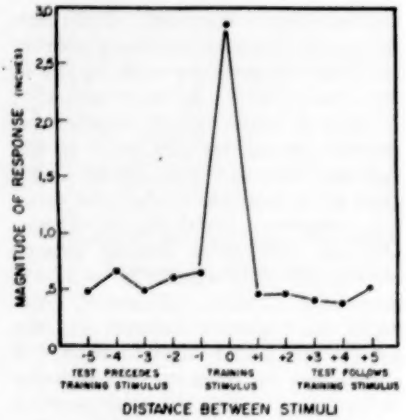

FIG. 4. Mean magnitude of response (inches of movement) as a function of the distance of the test stimuli (forward and backward) from the training stimulus in the original learning list.

Latency.—Measures of response latency were obtained for responses to the training stimulus during the testing period and for responses to the test stimuli. The scores used for analysis were the mean latencies of each *S*'s response to each of the stimuli. Analysis of variance revealed no significant differences among means as a function of the serial position of the training stimulus during original learning.

The mean latency of each *S*'s response to the training stimulus was also compared with the mean latency of his response to the test stimuli. The difference between these means was found to be highly significant ($t = 7.09$; $P < .001$). This difference, however, was just the reverse of that expected by our hypothesis (means were .40 sec. and 32 sec. to training and test stimuli, respec-

tively). This is also contrary to expectations based upon simple conditioning of autonomic responses. One might question, then, whether this is a "real" difference or just a one-in-a-thousand chance occurrence. Results reported by other investigators when *voluntary* responses were used as the conditioned (or learned) response indicate that this is not just a rare chance occurrence. Gibson (**8**) found that the average latency of generalized verbal responses to vibratory stimuli was less than that to the designated stimulus. She concluded that this difference was due to the inhibition of "long-latency false responses." Brown, Bilodeau, and Baron (**1**) found similar results when the task was to raise the finger from a reaction key when a designated light was turned on but to refrain from making the response when any other light on the same panel was turned on. Their data were highly unstable and differences were not statistically significant, but the trend seemed to support Gibson's findings. More recently, Eriksen (**7**) and Rosenbaum (**12**) have reported similar results.

Discussion

The finding that when the response can be controlled voluntarily latencies of responses to test stimuli are shorter than latencies to training stimuli thus seems to be a general one. Furthermore, it would appear to cast serious doubt upon the validity of the mediated generalization hypothesis. However, this does not necessarily follow; there is the possibility that this finding is an artifact of the experimental situation. We have already indicated that the mediated generalization process is probably an automatic process, not depending upon conscious use of the mediating material. Hence, all or most of the responses to test stimuli should be involuntary responses and should have relatively short latencies (making the reasonable assumption that latencies of voluntary responses are longer than latencies of involuntary responses). On the other hand, at least some of the responses to the training stimulus are voluntary, since *S* can wait until he is certain that the stimulus is the correct one and still make the response. Such voluntary responses may be expected to have unusually long latencies. In fact, long latency responses to the training stimulus when it immediately followed an erroneous response were very frequent in our data. These long latency voluntary responses to the training stimulus could only increase the mean, thus producing the observed effect.

At any rate, the fact remains that the frequency and magnitude data generally agree with expectations based upon the mediated generalization hypothesis. It seems necessary to assume that at least some of the erroneous responses are due to the direction and number of associations formed among the stimuli prior to training.

Summary

This experiment tested the hypothesis that erroneous responses which occur in a stimulus response situation are determined in part by the direction and number of associations which have been formed among the stimuli prior to training.

Seventy-two *Ss* were required to learn a list of six nonsense syllables by the method of anticipation to a criterion of eight successive correct trials. Then these *Ss* were divided into six groups of 12 *Ss* each. Each group was trained to make a motor response to a syllable which had occupied one of the six positions in the original learning list (a different position for each of the six groups). The *Ss* were then tested by four presentations of each of the other five syllables in the list, presentations being interspersed randomly among 35 additional training trials.

Frequency and magnitude of response to test stimuli tended to increase as the position of the training stimulus varied from the first to the last half of the original learning list, in agreement with the hypothesis. Also, significantly more generalization occurred to test stimuli which had immediately preceded than to those which had immediately followed the training stimulus in the original list.

Contrary to expectations based upon the mediated generalization hypothesis, the latencies of responses to test stimuli were shorter than those to the training stimulus. It was pointed out that this result casts serious doubt upon the mediated generalization hypothesis, but that there exists the possibility that this finding is due to factors inherent in the experimental situation.

References

1. Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.
2. Bugelski, B. R. A remote association explanation of the relative difficulty of learning nonsense syllables in a serial list. *J. exp. Psychol.*, 1950, **40**, 336–348.
3. Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.
4. Cason, H. Specific serial learning: a study of backward association. *J. exp. Psychol.*, 1926, **9**, 195–227.
5. Cason, H. Specific serial learning: a study of remote forward association. *J. exp. Psychol.*, 1926, **9**, 299–324.
6. Deese, J., & Kresse, F. H. An experimental analysis of the errors in rote serial learning. *J. exp. Psychol.*, 1952, **44**, 199–202.
7. Eriksen, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–565.
8. Gibson, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, **24**, 237–253.
9. Raskin, E., & Cook, S. W. The strength and direction of associations formed in the learning of nonsense syllables. *J. exp. Psychol.*, 1937, **20**, 381–395.
10. Razran, G. Semantic and phonetographic generalizations of salivary conditioning to verbal stimuli. *J. exp. Psychol.*, 1949, **39**, 642–652.
11. Razran, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.
12. Rosenbaum, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
13. Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.
14. Silver, C. A., & Meyer, D. R. Temporal factors in sensory preconditioning. *J. comp. physiol. Psychol.*, 1954, **47**, 57–59.
15. Wickens, D. D., & Briggs, G. E. Mediated stimulus generalization in sensory preconditioning. *J. exp. Psychol.*, 1951, **42**, 197–200.

(Received September 19, 1955)

Journal of Experimental Psychology
Vol. 52, No. 4, 1956

ROLE OF RESPONSE VARIABLES IN RECOGNITION AND IDENTIFICATION OF COMPLEX VISUAL FORMS [1]

HAROLD W. HAKE AND CHARLES W. ERIKSEN

The Johns Hopkins University

A recent article (6) describes the evidence which had been produced to show that practice in labeling a set of unfamiliar visual forms increases the ability of *Ss* subsequently to recognize each of the forms in the set. Although this fact seemed well established (**1, 4, 5, 7, 8**), there was no clear indication of the precise nature of the role played by the labels themselves. There was no reason to suggest that practice in associating the verbal responses with the stimulus forms had a significance beyond the mere fact that it constituted practice in discriminating among the forms.

It was concluded that the perceptual gain resulting from labeling practice may arise simply because the task requires *S* to search for the stimulus aspects by which the forms can be discriminated, and that this requirement can be imposed by any response set requiring *S* to make decisions about the forms. This appeared to hold even though *S* was using as few as two responses to reflect his discriminations among as many as 16 forms (6).

The possibility exists that these essentially negative results can be related to the fact that the verbal responses studied were "irrelevant" responses. That is to say, they were chosen to have a minimum of relevance to the stimulus forms and to have few private and uncontrolled associations and significances for *Ss* prior to the experiment. Letters of the alphabet were considered suitable because they are familiar to *Ss* but are not used habitually as identifying responses (6). The *Ss*, consequently, were not using verbal responses which they were accustomed to using in labeling or naming objects.

The experiment reported here studies systematically the effect of giving *Ss* practice in using alphabet letters as discriminating responses prior to learning to associate them with a standard set of unfamiliar visual forms. This variable was combined with that of the previous study, the number of letters which *Ss* were permitted to use in labeling a constant set of forms (6). This last variable is considered to be a critical one in determining the importance of the verbal responses used in labeling practice. Variation of the number of labels used tests the hypothesis that results of labeling practice are due to the specific associations formed during practice between unique responses and each of the forms.

Immediately following these paired-associate trials in which *Ss* associated their labels with the standard set of unfamiliar forms, *Ss* were required to recognize these forms, without naming them, as they appeared with other similarly constructed new forms. This latter situation provided a means of measuring the possible effects of verbal labeling practice independently of *S*'s ability to use the label sets.

[1] This research was done under Contract No. AF 33 (038)-22642, between the Institute for Cooperative Research, The Johns Hopkins University, and the Psychology Branch of the Aero Medical Laboratory, Directorate of Research, Wright Air Development Center.

METHOD

Procedure.—Except for *Ss* in the control groups who were given no initial experience in the use of the standard sets of labels, the procedure consisted of three tasks. In Task I, *Ss* in the experimental groups were given initial experience in applying the standard sets of labels to either a set of 12 familiar geometric forms or else a set of 12 air photographs of the city of Baltimore, Maryland. In Task II, all *Ss* labeled a standard set of unfamiliar complex visual forms. All *Ss* in the experimental groups used the same labels in Tasks I and II. Four sets of labels were used containing, respectively, 2, 4, 6, or 12 labels. In Task III all *Ss* saw the forms which had appeared in Task II presented together with 12 other forms of the same type. Subjects were required merely to indicate those forms which they thought they had seen in Task II.

Eight experimental groups of *Ss* were used representing the eight combinations of the four sizes of label sets with the two types of experience given in Task I. Four control groups, one for each of the four label sets, experienced only the procedures of Task II and Task III.

Each of the learning tasks, Task I and Task II, consisted of 120 trials in which each *S* saw each of the 12 forms 10 times. The forms appeared in a partially restricted random order which provided for the appearance of all 12 patterns in the first 12 trials, in the second 12 trials, etc. Successive appearances of the forms were spaced 10 sec. apart and each form was presented for 2 sec.

At the beginning of each trial a form was placed in view of *S*. Immediately following this *S* was required to assign one of the verbal labels to the form. Each *S* had a card in front of him on which was printed a list of labels which he could use. These consisted of letters of the alphabet chosen to have as little sequential connection as possible. After *S* on each trial had identified the form, *E* then pronounced the arbitrarily predetermined correct label for that form. The task of *S* in the first two tasks, then, was to learn which of the labels from his list could be correctly applied to each of the patterns.

In each of the 12 groups one-half of the *Ss* were required to learn an arbitrary set of label-form assignments which was different from that required of the remaining *Ss*. This variation in the nature of the label-form assignments produced no significant effects in any of the three tasks, and, for this reason, will not be discussed again.

There were six trials in Task III, the recognition task. On each trial *S* was given an 8½ × 11-in. sheet on which appeared the 12 standard complex forms he had labeled in Task II together with 12 other similar forms that he had not seen previously. On each of the first two trials the standard forms were rotated 90° to the left or right. On the Trials 3 and 4 the forms were rotated 180°. On Trials 5 and 6 all of the standard forms appeared in the positions in which they had appeared when *S* labeled them in Task II. The systematic rotation of the standard forms was designed both to make the task of recognition somewhat more difficult and to require *S* to generalize information which he had gained about the forms during the second task. The 12 new unfamiliar forms introduced in Task III were also rotated systematically from trial to trial in an attempt to prevent *S*'s learning about the characteristics of these new forms. Each of the 12 standard and 12 unfamiliar forms occupied one of the 24 cells, 1 in. square, formed by the intersection of four vertical columns and six horizontal rows. Except for being rotated on some of the sheets, each standard form appeared in the same size and color in Task II and Task III.

Each sheet was presented to *S* for 1 min. and during this time he was required to check those forms that he thought were the ones he had labeled in Task II. He was told in advance of each trial that the forms could appear in rotated positions and that each sheet contained 12 standard forms and 12 unfamiliar forms.

The procedures of all three tasks were completed within a single 1-hr. session.

Apparatus.—Each of the forms used in the first two tasks was mounted within a 1-in. square area in the center of 3 × 5-in. index card. The over-all dimensions of the forms varied somewhat; the smallest dimension was about ¾ in., the largest almost 1 in.

The 12 standard and the 12 unfamiliar nonsense forms which *Ss* saw in Task II and Task III were "nonsense" closed-line drawings constructed by following certain rules chosen by a partially restricted chance selection. These forms are shown in Fig. 1 which represents one of the sheets shown to *Ss* in Task III. The air photo forms were small segments, about 1-in. square, of a series of large air photographs of the city of Baltimore published by the Baltimore *Sun* over a period of several months in 1954. The segments chosen were mainly top views of the geometrical configuration formed by typical blocks of Baltimore "row" houses. The geometric forms included a square, circle, triangle, cross, diamond, a rectangle with a long vertical dimension, two horizontal bars, a parallelogram inclined to the right, an ellipse with a long vertical dimension, an ellipse with a long

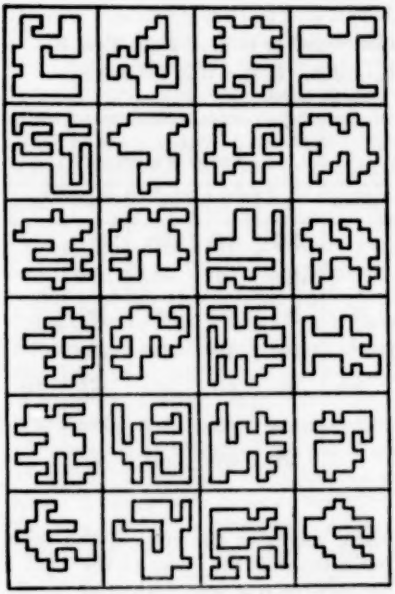

FIG. 1. The nonsense forms used in the experiment in Task II and Task III. This matrix of forms is identical, except for scale, to one of the sheets presented to *Ss* in Task III.

horizontal dimension, a six-pointed star, and a semicircle.

Subjects.—A total of 120 *Ss* was used, 10 in each of the 12 groups. They were male undergraduate students at The Johns Hopkins University who were run through the experimental procedures one at a time.

RESULTS

The data considered first are the proportions of correct labeling responses achieved in the first two tasks. These require correction for chance effects in both tasks. Subjects were using different numbers of labels in identifying the 12 forms and thus could achieve correct labeling responses by means of the chance expectancies associated with the number of labels used. Consequently, the performance of each *S* was corrected for chance by the formula used by Bricker in a situation similar to ours (**3**).[2]

Task I.—In the left half of Table 1 we present the average proportions of correct labeling responses achieved by the 10 *Ss* in each of the eight experimental groups during the 120 trials of labeling practice The proportion of correct responses was computed for each subject individually using the correction formula described by Bricker (**3**) The groups are arranged in columns according to the number of labels used in Task I and in rows according to the forms being labeled.

An analysis of variance of the results of Task I indicated significant effects attributable to the number of labels being used and the type of stimulus materials labeled.[3] Subjects labeling the geometric forms had more success in learning to associate the labels with the forms independent of the number of labels used, than did *Ss* learning to associate the labels with the air photo forms ($P < .001$).

The decreasing accuracy of labeling, corrected for chance, with increasing number of response categories was an unexpected result ($P < .05$). We had not found this result in a previous study (**6**); nor had Bricker (**3**). The analysis of variance term for the interaction of number of labels used and type of stimulus material being labeled was not significant, indicating

[2] The proportion correct was computed from a number correct score which had been corrected for chance by the use of:

$$R_c = R - W(p_r/p_w),$$

where R_c is the corrected number right, R is the raw number right, W is the number wrong, p_r is the probability that a correct response will occur by chance, and p_w is the probability that a wrong response could occur by chance.

[3] The group variances in this analysis and those in all subsequent analyses can be assumed homogeneous. Prior to the analysis of variance in each case, group variances were tested for heterogeneity by Bartlett's test (**2**).

TABLE 1

PROPORTION OF CORRECT LABELING RESPONSES ACHIEVED IN TASK I AND TASK II

(Row designation refers to Task I experience of *Ss*. In Task II all *Ss* labeled the standard set of nonsense forms)

Task I Experience	Task I Number of Labels Used					Task II Number of Labels Used				
	2	4	6	12	Means	2	4	6	12	Means
None						.330	.411	.358	.247	.337
						.620	.651	.602	.396	
Labeled air photos	*M* = .593	.602	.500	.507	.551	.383	.387	.438	.389	.399
	SD = .598	.542	.522	.675		.809	.599	.551	.809	
Labeled geometric forms	.860	.678	.850	.689	.769	.387	.431	.690	.553	.515
	.558	.764	.333	.916		1.034	.665	.568	.806	
Means	.727	.640	.675	.598		.367	.410	.495	.396	

that the relation between accuracy and number of labels used was not affected significantly by the type of material being labeled.

Task II.—Average performance of *Ss* in each of the 12 groups in Task II is presented in the right half of Table 1. In the first row are the four control groups of *Ss* who did not participate in Task I. In the second row are the four groups of *Ss* who labeled the air photo forms in Task I, and in the third row are the four groups of *Ss* who labeled the geometric forms in the first task. In Task II all *Ss* labeled the standard set of nonsense forms.

An analysis of variance of the data of Task II indicated that the effect of the stimulus materials labeled in Task I was significantly large ($P < .05$). Subjects who had labeled the air photo forms in the first task did significantly less well than did *Ss* who had labeled the geometric forms prior to Task II. Subjects in both of these groups did better in the second task than *Ss* who had not used their labels previously. Although consideration of the group mean scores in the cells of the table indicates that the effect of the Task I procedures was less noticeable in the case of *Ss* using the smaller numbers of labels, there was no evidence of a significant interaction between the number of labels used in Task II and the type of Task I experience. The effect of previous experience with the labels held for all of the label sets used in the experiment. On the other hand, the effect of the number of labels used in the tasks, independent of the type of stimulus forms labeled in the first task, was not significantly large. This latter finding suggests that the effect of practice in the use of labels was most pronounced for *Ss* using larger numbers of labels. These *Ss* had achieved significantly fewer correct labeling responses in Task I than did *Ss* using smaller numbers of labels.

Task III.—In the third task *Ss* were required merely to recognize the forms labeled in Task II. The data averaged over the six trials of this task are presented in two forms: in the left half of Table 2 is the average

number of forms identified as familiar; in the right half of the table is shown the relative accuracy with which the forms were identified as familiar. The latter score was computed by dividing the number of standard forms which were called familiar by the total number of all forms which *Ss* thought they recognized. Each value in the table represents the average performance of the 10 *Ss* in each group.

An analysis of variance performed upon each half of Table 2 separately revealed no clear-cut effects of either the number of labels used or the kind of experience which *Ss* had with their labels prior to Task III. The effect of the number of labels used in Task II upon the number of forms recognized in Task III, without regard to the accuracy of such recognitions, was significant ($P < .05$). This result is ambiguous, however, because the interaction of this variable with the type of forms which *Ss* had labeled prior to Task II was highly significant ($P < .025$). This tendency for *Ss* in some of the groups to make significantly more recognition responses does not appear related in a simple way to the main variables of the experiment.

The only significant effect shown by the analysis of the data in the right half of Table 2 was also due to the interaction effect ($P < .05$). This significant interaction term could be identified with unusually high relative accuracy on the part of *Ss* in four of the groups. This result, again, does not appear to be related in any systematic way to the group results of Task I or Task II. In general, the performances on Task III do not appear to reflect in any simple way the effects of variables introduced in the first two tasks.[4]

[4] The results were analyzed separately for Trials 1 and 2, Trials 3 and 4, and for Trials 5 and 6. In general, this further breakdown provided results similar to those reported for all trials taken together. The only additional information added was that recognition of familiar figures in the upright position on Trials 5 and 6 was significantly better than in Trials 1 through 4.

TABLE 2

NUMBER OF RECOGNITION RESPONSES GIVEN IN TASK III AND THE PROPORTION OF THESE RESPONSES WHICH WERE CORRECTLY GIVEN TO FAMILIAR FORMS

Task I Experience	Number of Recognition Responses					Proportion Correct Recognition Responses				
	Number of Labels Used Previously					Number of Labels Used Previously				
	2	4	6	12	Means	2	4	6	12	Means
None	*M* = 64.4	63.8	46.5	63.7	59.6	.641	.647	.738	.665	.673
	SD = 31.08	17.09	42.90	32.40		.210	.265	.358	.207	
Labeled air photos	53.8	64.8	58.4	61.6	59.6	.765	.693	.668	.677	.701
	46.08	26.94	31.11	36.22		.401	.370	.348	.402	
Labeled geometric forms	67.3	62.7	60.7	52.4	60.8	.645	.654	.777	.719	.699
	20.93	23.24	27.24	48.58		.385	.377	.251	.359	
Means	61.8	63.8	55.2	59.2		.684	.665	.728	.687	

TABLE 3

WITHIN-SUBJECT CORRELATION COEFFICIENTS COMPUTED FROM NUMBER OF TIMES EACH FORM WAS CORRECTLY LABELED IN TASK II AND CORRECTLY RECOGNIZED IN TASK III

Task I Experience	Number of Labels Used				Mean Correlation
	2	4	6	12	
None	.218*	.213	.206	.141	.193**
Labeled air photos	.017	.233	.402	.350	.265
Labeled geometric forms	− .007	.207	.399	.422	.251

* Computed from pooled variates from 10 *Ss* in each group, each deviation involved being measured from its own subject mean.
** Computed from pooled variates from 30 *Ss* who had same Task I experience.

Within-S correlations.—Another possible effect of the main variables of this experiment was upon the degree to which performance *within Ss* during Task II was related to performance during the recognition task. If the act of learning to associate labels with each form affected the degree to which *S* learned about the discriminable aspects of these forms, we could expect that *S* would most reliably recognize in the third task those forms that he had labeled most successfully in the second task. On the other hand, if the degree to which *S* learned about the forms were independent of the success with which he labeled each form, we could expect insignificant within-*S* correlations between performances of each *S* on the two tasks. The previous experiment indicated that this relationship accounted for a significant, but minor, part of the variance within *Ss* in learning new labels for a set of stimulus forms with which they had previously learned to associate another set of labels (6).

Average within-*S* correlations for each of the present groups are shown in Table 3 where the groups are ordered by rows and columns as before. These coefficients are based upon the number of times that a given stimulus was labeled correctly in Task II and recognized correctly in Task III by each *S*. The coefficients were computed from pooled variates from the 10 *Ss* in each group, each deviation involved being measured from its own *S* mean (9).

Within-*S* correlations are significantly affected by practice in the use of sets of labels, and the specific effect found is significantly related to the number of labels in the set being used. Within those *Ss* who labeled the air photo forms or the geometric forms in Task I there is a statistically significant increasing degree of correlation between performance in Task II and Task III with increasing number of verbal responses ($P < .05$).[5] This is not true of *Ss* who had no

[5] These significant changes in the magnitude of correlations are not due simply to systematic changes in the magnitude of the variances of the predictor or predicted variates. The within-*S* variances were homogeneous among all groups of *Ss* who experienced the same labeling practice in Task I.

prior experience with their labels before using them to identify the standard nonsense forms in Task II. Within these *Ss* the correlations are of the same order of magnitude consistently found within the *Ss* of the previous study (**6**).

Discussion

There appear to be three outstanding results of the present experiment. The first emphasizes aspects of verbal labeling performance which previously have been neglected. These are aspects of verbal labeling skill which are specific to the responses used rather than to the stimulus forms being labeled. It should be expected, for example, that the changes which occur while *Ss* use verbal labels for the first time in attempting to label a set of forms should include increases in the skill with which the responses are used as well as progressive differentiation of the forms. The *Ss* during such practice must learn to use their responses discriminatively as well as learn to discriminate among the forms. Since one can reasonably expect response sets differing in the number of responses in each, to differ in the extent to which *Ss* can quickly acquire skill in their use, there should exist aspects of verbal labeling performance which are characteristic of the use of a particular number of responses, independent of the type of stimulus materials being labeled.

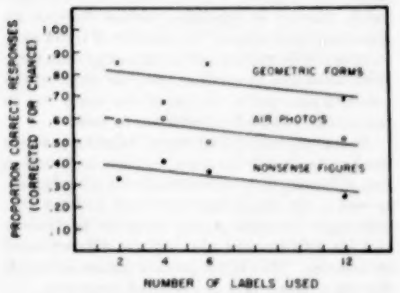

Fig. 2. Proportions of correct labeling responses achieved plotted as a function of the number of different labeling responses being used. The lines fitted to the data represent a least-squares fit to all the points adjusted for the mean labeling proficiency achieved in the case of each type of stimulus form.

Evidence supporting this suggestion is supplied by Fig. 2 where is plotted the proportion of correct labeling responses achieved by *Ss* using sets of labels for the first time to label forms differing in initial familiarity. The plotted points are the Task I performance of *Ss* who labeled the air photo and geometric forms in that task and the Task II performances of *Ss* who had no Task I experience.

An analysis of the relationship between the proportion of correct labeling responses and the number of labels used indicated that this is a significant linear trend decreasing as the number of labels used increases ($P < .05$). More importantly, this trend can be assumed to be the same for all sets of stimulus forms, indicating that an aspect of labeling skill can be independent of the stimulus sets being labeled and specific to the number of labels being used. When larger and larger response sets are used, *Ss* have increasing difficulty in learning to associate the responses with the stimulus forms; and this same relationship holds for sets of stimulus forms differing in initial familiarity. The lines fitted to the data represent a least-squares fit to all the points adjusted for the mean labeling proficiency achieved in the case of each type of stimulus form (**9**).

A second result of the experiment is the demonstration that increased skill in the use of verbal responses need not be accompanied by increases in discriminative skill specific to the stimulus forms labeled. That is, increases in skill in pairing responses with the forms need not be accompanied by increased stimulus differentiation. The effects of practice in the use of sets of labels previous to Task II did significantly increase skill in labeling the nonsense forms in Task II. This enhanced ability to correctly associate the labels with the forms, however, did not increase the ability of *Ss*, considered in groups, to

recognize these forms when presented with others in Task III. Thus, it is evident that variables exist which can affect the efficiency of labeling practice without significantly affecting at least one type of perceptual discriminative skill which can result from that practice.

This negative result contrasts interestingly with the third result considered here. This is the evidence produced to indicate that *within* *S*s previous practice in the use of labels prior to Task II could affect the success with which each form was recognized in the third task. Providing *S*s are given previous practice in the use of labels, the degree of success achieved in subsequently using these labels to identify a new set of forms can become a good indicator of the degree to which they learn about an aspect of the forms while making these verbal identifications. That is, the perceptual gain, of the sort tested by Task III, becomes more reliably tied to a measure of previous labeling performance.

This suggests that when unfamiliar labels are used without prior opportunity to acquire skills specific to the labels, correct labeling responses have little significance. Subjects learn about the forms whether or not correct labeling responses are achieved. With previous practice in the use of the labels, however, it appears that what *S*s learn about the forms becomes more dependent upon the *correctness* of labeling responses.

Since the effect of practice upon the within-*S* correlations was significantly related to the number of labels being used in the labeling tasks, it appears reasonable to suppose that familiar labeling responses may have a *denotative* function for *S*s. They may be useful to *S*s in retaining from trial to trial the identity of some of the forms and certain information associated with them such as which forms had been correctly labeled on a previous trial, which critical differentia appear in which forms, which forms have certain differentiable features in common, etc. If the labels had such a denotative function, it should be expected that *S*s using the smaller numbers of labels would find them less useful in this way. This appears to be the case. The effect of previous experience in the use of just two labels upon the within-*S* correlations, for example, is to reduce them to zero.

These results taken together suggest two possible effects which can be produced by verbal labeling practice. One of these is a process involving both increased differentiation of the set of stimulus forms and increased skill in the use of verbal labels. The other process, the appearance of which depends upon the use of larger sets of practiced responses, may be the denotative use of labels to organize and identify the discriminable aspects of stimulation which are differentiated by practice.

Summary

Subjects were given prior practice in the use of sets of irrelevant labeling responses before learning to associate them with a set of unfamiliar nonsense forms. Although previous practice in the use of the labels did significantly increase the number of correct labeling responses achieved by *S*s in labeling the nonsense forms, it did not increase their ability to recognize the forms later when seen together with new forms of similar construction.

Previous practice did affect the size of within-*S* correlations. With larger sets of practiced labels, the number of times each form was correctly labeled in labeling practice became an important indicator of the number of times each form would be recognized in the recognition task. With label sets as small as two, the effect of the previous practice in the use of the labels was to reduce the within-*S* correlations to zero.

It was suggested that verbal labeling practice can have a double function. It has the function first of forcing *S*s to differentiate the stimulus set, as well as the set of responses used, and can provide also a *denotative* process whereby *S*s organize and identify the stimulus aspects differentiated by practice. The latter process occurs only with the use of larger sets of practiced responses.

References

1. Arnoult, M. D. Transfer of predifferentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45**, 401–409.

2. Bartlett, M. S. Some examples of statistical methods of research in agriculture and applied biology. *Suppl. J. Roy. statist. Soc.*, 1937, **4**, 137–183.
3. Bricker, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49**, 73–81.
4. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
5. Goss, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46**, 419–428.
6. Hake, H. W., & Eriksen, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50**, 161–167.
7. Robinson, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–115.
8. Rossman, I. L., & Goss, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.
9. Snedecor, G. W. *Statistical methods.* (3rd Ed.) Ames, Iowa: Collegiate Press, 1940.

(Received September 9, 1955)

Journal of Experimental Psychology
Vol. 52, No. 4, 1956

THE INTERACTION OF SEVERAL VARIABLES IN LATENT LEARNING[1]

OTELLO DESIDERATO[2]

New York University

Survey of the literature reveals that most latent learning experiments have represented attempts to prove or disprove a specific theoretical formulation. Instead of adding to the number of "crucial" tests of existing theories, it was felt to be more desirable at present to explore the effects of several variables commonly associated with latent learning experiments.

The present paper deals with a situation which Thistlethwaite (**14**) has classified among the Type II latent learning experiments: *S*s are permitted to explore an empty maze; after these exploratory periods, a relevant goal object is introduced into one portion of the maze, and the test trial is given. Both Seward (**12**) and the author (**5**) have found positive evidence of latent learning in this type of experimental situation. The present experiment is exploratory in nature in that it investigates the gross effects and interactions of five variables on the amount of latent learning demonstrated in a Type II situation. The variables, each of which varied in two ways, were:

Straight-alley pretraining.—In view of several reports in the literature which suggest the positive effect of pre-experimental training (**2, 3, 4, 7, 8, 15**), it was hypothesized that a period of straight-alley pretraining prior to maze exploration training would enhance latent learning.

Visual cues.—The findings of Seward (**12**), Honzik and Tolman (**6**), and the present author (**5**) suggested that rats will utilize visual cues received during their transfer from the prefeeding location in the maze to the start box on the test trials. Hence, it was hypothesized that *S*s permitted to receive such cues would perform better on the test trials.

Type of exploration.—The incentive function of cage mates in maze learning has been mentioned in previous studies (**9, 11, 13**). It appeared legitimate, therefore, to explore the effects of "social reinforcement" in the present situation, assuming that some reward is operating when cage mates explore an empty maze together. The precise mechanism whereby such reinforcement affects learning was left unspecified. Thus, it was predicted that *S*s exploring a maze together with cage mates would show more latent learning than individual explorers.

Activity drive.—It is general practice to restrict rats to the tiny compartments of carrying cages while they await their experimental trials. It was hypothesized that severe *restriction* of skeletal activity just prior to each maze exploration period would increase the need for activity, so that subsequent **T** maze exploratory responses per se would be reinforced. Another factor that might be influenced by the amount of activity drive is that more active rats would run faster and have less time intervening between the choice point and goal box. This should enhance learning by increasing the temporal contiguity of critical maze stimuli. On the other hand, rats forced to tread inside an activity wheel were predicted to show

[1] This study is based upon a dissertation submitted in 1953 to the Graduate School of Arts and Science, New York University, in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express his gratitude to Prof. Howard H. Kendler for his generous advice and encouragement.

[2] Now at Adelphia College.

poorer learning, since they would start their exploratory periods with a decreased need for activity.

Vestibular turn.—In a pilot study (**5**) no latent learning was found when each *S*, held in *E*'s hands, was given a complete turn in space during its transfer from prefeeding goal box to start box on the test trial. This finding suggested that the constellation of choice point cues to which the successful rat responds on the test trial probably includes a characteristic pattern of vestibular excitation aroused by the customary transfer movement. It was thus hypothesized that sufficient variation in the pattern of this excitation should alter the compound of relevant choice-point stimuli, and thus reduce the probability of emission of the appropriate response. It was assumed that the 360° turn interpolated within the transfer process would sufficiently modify the pattern of internal stimulation to measurably reduce the degree of learning manifested. Hence, *S*s receiving this rotational movement during their transfer from prefeeding goal box to start box for the test trials were expected to show impairment of latent learning.

Method

Subjects.—Ninety-six experimentally naive male albino rats of the Wistar strain were used, their ages ranging between 90 and 105 days at the beginning of the experiment.[3]

Apparatus.—The single-unit T maze was essentially similar to that employed by Seward (**12**). The white goal box on the right and the black one on the left were placed at right angles to the maze arms. The rest of the maze was painted flat gray. The length of the stem was 4 ft.; the arms, 8 ft. over all. Alleys and goal boxes were 4 in. wide. The walls were 5½ in. high throughout, and the roof was covered with ¼-in. mesh hardware cloth. The entire structure was enclosed within a cheesecloth screen suspended over a string frame running 18 in. above the roof of the maze. Four 15-w. lamps, mounted in upturned coffee tins to serve as reflectors, were suspended from the ceiling in fixed positions directly on top of the cheesecloth screen: above both goal boxes, above the choice point, and above the start box in the maze stem. Light sheet-metal guillotine doors were located at the start box in the maze stem, on either side of the choice point, and at the entrance to each goal box. The far end of each goal box was also provided with a guillotine door so as to permit the rat exit from the goal box into an adjacent carrying box.

A transfer box was used on test trials to carry *S*s from the goal box to the start box, in order to preclude the reception of visual cues. This transfer box was 10 in. long, 4 in. wide, and 5½ in. high, and was completely enclosed except on one side, where entry or exit was afforded by raising a sliding metal panel.

The straight alley resembled the maze in color and inside width. Over-all length was 68 in. At one end, black or white goal boxes, resembling T-maze goal boxes, could be interchanged.

Restriction chambers were fashioned by reducing the size of standard carrying cage compartments until they measured 7 in. long, 4 in. high, and 2¾ in. wide, leaving each *S* barely enough room in which to turn around.

A power-driven activity wheel was constructed having several compartments, each 5 in. wide, separated by plywood partitions. The cylinder measured 16 in. in diameter and was covered with ½-in. wire mesh. A 1/75-hp AC motor rotated the wheel at a constant speed of 5 rpm, forcing the rat within to move a distance of 50 in. per revolution.

Procedure.—All *S*s were started on a 21-hr. hunger cycle 10 days before exploratory training began. Procedure common for all *S*s was to provide 4 days of exploration in the empty T maze, 40 min. per day. Each goal box contained an empty food dish. On the fifth day, one *S* at a time received additional 3 min. of maze exploration. One-half hour later, *S* was placed through the roof into one goal box and was allowed to eat wet mash for 1 min. It was then immediately brought to the start box where it was given a free-choice trial. Either response led to an unbaited goal box, containing only an empty food dish. The *S* was detained in the goal box of its choice for 30 sec. Ten such trials were given on each of two consecutive days, with 15 min. between trials. The prefeeding period was reduced to 30 sec. on every trial beyond the first.

The number of consecutive correct responses to the prefeeding goal box was the response

[3] In order to estimate whether these rats were too young to learn the task, the main experiment of Seward was reproduced, using *S*s only 60 days old. Evidence of latent learning was found at the .065 level of significance. It was concluded that rats 30 to 45 days older were sufficiently mature.

measure recorded. The assumption underlying this extinction measure was that the stronger the learning of the location of the black and white goal boxes, the greater would be the tendency to return to that goal box in which prefeeding occurred before every trial. That is, if *S* had learned that a white goal box was on the right side, and if it is then prefed in this goal box before every trial, it should continue to visit the white goal box longer than an *S* with inferior knowledge of the goal boxes' locations, even though the correct response was never directly followed by food reward.

Conditions.—The experimental groups were arranged in a 2^5 complete factorial design, with three replications per cell.

Half the *S*s received straight-alley pretraining for six days before the commencement of exploratory training (Group P). The first three days, these rats were permitted to explore the alley for two 10-min. periods, once running to a black goal box, once to a white goal box. The last three days, every *S* was given two runs to the goal boxes, once to W, once to B. The remaining half of the *S*s were simply handled 3 min. daily for the six days (Group NP).

Half the *S*s explored the straight-alley and T maze in groups of three cage mates at a time (Group SOC). The other half explored individually (Group IND).

Half the *S*s were confined within small restriction chambers for 1 hr. before each maze exploration session (Group RE). The other half was forced to tread within a motor-driven activity wheel for 15 min. before each maze exploration period (Group AC).

On the test trials, half the *S*s were transferred from prefeeding goal box to start box within the enclosed transfer box, eliminating visual cues (Group NC). The remaining *S*s were transferred by hand, being lifted out through the roof of the goal box, in full view of the apparatus and room (Group C). During this transfer, half the *S*s were given a full 360° turn in space, coming to rest at the start box (Group V). For the other half, this "vestibular" turn was omitted (Group NV).

Results

In order to satisfy the requirement of homogeneity, the $\sqrt{X + .5}$ transformation was applied (1). Table 1 shows the means and *SD*s of the raw and transformed data for each condition. Table 2 summarizes an analysis of variance of the transformed scores. The within-cells error term was used for evaluating all effects. This table reveals that the obtained difference between the two groups composing the Visual Cues variable is accepted as a chance fluctuation around the population mean. Since this variable does not enter into any interactions, the hypothesis that visual cues received during the transfer from prefeeding goal box to starting box affect performance on test trials is rejected.

TABLE 1

Means and *SD*s of Successive Correct Responses Before Occurrence of the First Error

Condition	Original		Transformed	
	Mean	*SD*	Mean	*SD*
Social exploration	2.92	3.31	1.65	.84
Individual exploration	4.44	4.48	2.00	.97
Visual cues	3.98	3.74	1.93	.88
No visual cues	3.38	4.24	1.72	.95
Alley pretraining	3.85	3.67	1.88	.91
No alley pretraining	3.50	4.32	1.77	.93
Vestibular turn	3.38	3.09	1.80	.80
No vestibular turn	3.98	4.74	1.85	1.03
Enforced activity	2.38	2.76	1.51	.77
Enforced restriction	4.98	4.60	2.14	.95

TABLE 2

Summary of Analysis of Variance of Transformed Latent Learning Scores

Source	*df*	*MS*	*F*
Activity drive (A)	1	9.56	14.60***
Alley pretraining (P)	1	.29	
Vestibular turn (V)	1	.06	
Visual cues (C)	1	.97	1.49
Type of exploration (T)	1	3.02	4.61*
A × V	1	4.90	7.48**
P × V	1	3.87	5.91*
P × T	1	3.25	4.96*
All other interactions combined	23	.60	
Within cells	64	.65	

* $P = .05$.
** $P = .01$.
*** $P = .001$.

Although the Activity variable is shown in Table 2 to be significant at the .001 level, this simple effect is accounted for by the magnitude of the significant Activity × Vestibular Turn interaction, which indicates that the superiority of restricted *Ss* over activity *Ss* emerges only when the vestibular turn is omitted.

Table 2 also reveals that the generality of the Type of Exploration main effect is conditioned by the Alley Pretraining × Type of Exploration interaction, indicating that the superiority of individual explorers over social explorers obtains only when alley pretraining is omitted.

Results of the Alley Pretraining × Vestibular Turn interaction show that alley pretrained *Ss* performed better than rats lacking this pretraining, when a rotational turn was given during the transfer from prefeeding goal box to the starting box on the test trials. If the turn was omitted, no significant difference was found between alley and no-alley pretraining groups.

To determine whether some experimental variables (e.g., enforced restriction or activity) operated differentially to enhance or depress the general level of activity during the T-maze exploration phase, analysis of variance was applied to the total number of empty goal-box visitations made by each *S* during the exploratory periods. The emergence of significant differences between conditions would suggest that differences on test trials might be related to different amounts of "practice" with the black and white goal boxes during the "latent" phase. Table 3 shows means and *SDs* for goal box visitations for *Ss* under different conditions, during T-maze exploration sessions. Table 4 summarizes an analysis of variance of goal-box visitation counts, after a square-root transformation. Reference to this table shows that *F* ratios for the Activity Drive main effect and for a number of interactions are significant.

To determine whether latent learning scores reflected differences in the frequency of goal-box visitations, i.e., experience with black and white goal boxes, an analysis of covariance was performed. Transformed goal-box visitation counts constituted the control variable;

TABLE 3

MEANS AND *SDs* OF BLACK AND WHITE EMPTY GOAL-BOX VISITATIONS DURING T-MAZE EXPLORATION PERIOD

Condition	Mean	*SD*
Social exploration	97.17	30.98
Individual exploration	86.35	26.72
Visual cues	95.13	29.30
No visual cues	88.40	29.18
Alley pretraining	95.33	30.94
No alley pretraining	88.19	27.39
Vestibular turn	96.54	28.98
No vestibular turn	86.98	29.10
Enforced activity	82.90	21.90
Enforced restriction	100.63	33.10

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF TRANSFORMED GOAL-BOX VISITATION COUNTS

Source	*df*	*MS*	*F*
Activity Drive (A)	1	19.52	11.38**
Alley pretraining (P)	1	2.45	1.43
Vestibular turn (T)	1	5.82	3.39
Visual cues (C)	1	3.72	2.17
Type of exploration (T)	1	5.96	3.47
A × P	1	10.43	6.08*
A × P × V	1	7.64	4.46*
V × C × T	1	7.68	4.47*
A × P × V × C × T	1	9.12	5.32*
All other interactions combined	22	1.65	
Within cells	64	1.72	

* $P = .05$.
** $P = .01$.

TABLE 5

SUMMARY OF COVARIANCE ANALYSIS BETWEEN TRANSFORMED LATENT LEARNING SCORES AND GOAL-BOX VISITATION COUNTS

Source	*df*	*MS*	*F*
Activity drive (A)	1	6.30	9.69*
Alley pretraining (P)	1	.16	
Vestibular turn (V)	1	.18	
Visual cues (C)	1	.65	1.00
Type of exploration (T)	1	3.57	5.49*
A × V	1	5.10	7.85**
P × V	1	3.73	5.74*
P × T	1	3.24	4.98*
All other interactions combined	23	.60	
Within cells	63	.65	

* $P = .05$.
** $P = .01$.

transformed test scores were used as the criterion variable.

Table 5 summarizes the results of this covariance. This table reveals that the same two main effects and the same three simple interactions which were significant in the variance analysis of the test trials (Table 2) still yield significant *F* ratios. These results indicate that even if the total number of goal-box visitations is statistically equated, the same variables will still interact in the same fashion. Thus, test trial performance cannot be traced directly to the number of goal-box experiences these variables induce.

DISCUSSION

The failure of the Visual Cues variable to yield a significant difference indicates that visual cues received during the transfer from prefeeding goal box to start box do not appreciably affect test trial performance. This result appears to conflict with the finding of Honzik and Tolman (6) that vision of the apparatus *in its relation to the room* was essential in order for their *S*s to avoid the shock side of a T maze. This discrepancy may be accounted for by noting that, in the present study, the cheesecloth screen may have sufficiently "camouflaged" the maze and its parts so that optimal conditions were lacking for responding to the maze parts in relation to the room. Moreover, the maze in this study was much larger, and the room was probably considerably darker, than the one used by Honzik and Tolman.

The superior learning of restricted *S*s over enforced activity *S*s (when the "vestibular turn" was omitted) has implications for the conduct of latent learning experiments and for further research on the relationship between activity drive and learning. It should be of interest to researchers that the forced inactivity of rats restricted to carrying compartments appears to considerably increase their general activity level when they are turned loose to explore the empty maze. Since their superior learning is not related to the number of times they actually enter discriminably different goal boxes, one may speculate that the heightened activity level operated to reduce the temporal gaps between choice point and goal boxes. Following MacCorquodale and Meehl s formulation (**10**), this would tend to increase the strength of expectancies by reducing the temporal separations between $S_1 \rightarrow R_1 \rightarrow S_2$. Another possibility is that enforced activity or restriction differentially affects the condition of the animal at the time it enters the goal box, making restricted *S*s more responsive to relevant goal box cues.

Since the vestibular turn was introduced just before each test trial (i.e., at the completion of exploratory training), the effect of this variable is one of interference upon what has already been learned. The perplexing datum is that this interference operates only on restricted rats. One hypothesis is that restricted *S*s learn to become more responsive to internal and external stimulation, by virtue of the potentiality of reinforcement for responding to all classes of stimulation, created by increases in the need for activity. One possibility is that the vestibular turn

would introduce novel patterns of internal stimulation which might interfere with the making of fine choice-point discriminations on the part of the more responsive restricted animals.

For *S*s lacking straight-alley pretraining, it was noted that individual explorers were superior to social explorers. A reasonable explanation is that cage mates exploring a T maze together spend considerable time responding to each other, rather than to differential features of the maze environment. Thus, the reception of stimulation necessary for learning the maze is interfered with by stimuli resulting from the behavioral interaction among the *S*s. Observation of social and individual explorers tends to support this explanation. In general, individual explorers display smooth, uninterrupted traversal over units of the maze, as from choice point to goal box. Social explorers spend much of the exploratory period chasing and licking each other, tumbling and climbing over one another, and in similar types of "playful" behavior. This was true especially at the beginning of each exploratory period, and more so during the first two periods than for the last two.

The failure to find a significant difference between individual and social explorers when the conditions included straight-alley pretraining suggests the hypothesis that alley pretraining may provide social explorers with an opportunity for some extinction of "responses to each other." After a total of one hour in the straight alley, social explorers may begin the exploratory phase of the experiment with reduced tendencies to respond to each other. Alley pretrained social explorers would thus be better prepared for learning to respond to the physical features of the T maze than social explorers lacking such pretraining. The latter would begin T-maze exploration with a stronger tendency to make social responses to cage-mate stimuli. The important point, however, is that apparently the strength of latent learning manifested is not independent of the number of *S*s exploring at one time nor of their pretraining experiences.

Additional effects of alley pretraining are suggested by the Pretraining × Vestibular Turn interaction. If the hypothesis that pretraining permits extinction of interanimal responses is broadened to include extinction of responses to internal cues, then by the beginning of exploratory training, alley pretrained *S*s should be responding chiefly to relevant maze stimuli. If such a hypothesis is correct, pretrained *S*s should be less sensitive to the interfering effects of the rotational turn on test trials. This would tend to explain the superiority of alley pretrained over nonpretrained *S*s, when both receive the vestibular turn. Indirect test of the hypothesis of response extinction to vestibular stimuli during alley pretraining can be made by varying the amount of pretraining for different groups, and subjecting all *S*s to the rotational movement.

The above interpretations are of course, all tentative, and require, if their validity is to be established, a large amount of additional evidence. For the present it may be appropriate to point out that much of the research in the area of such complicated phenomena as latent learning has been severely restricted to only a few variables. If we are to fully understand these phenomena, we must discover how other variables influence this behavior. The point of this research was to take a step in that direction.

Summary

Ninety-six male albino rats explored an empty T maze. One goal box was painted black, the other white. Before each test trial, each *S* was placed in one goal box, through the roof, where it found food for the first time. Free-choice test trials were given until the *S* made the first incorrect response. No test-trial response was reinforced. Analysis of variance indicated that rats subjected to restriction of activity prior to maze exploration showed superior learning to that of rats subjected to forced activity. The superiority of restricted rats was impaired if they were given a rotational turn in space during their transfer from goal box to start box on test trials. Individual explorers

were superior to rats exploring in the company of cage mates, but straight-alley pretraining eliminated this difference. Rats lacking alley pretraining suffered greater impairment of performance than pretrained rats when subjected to a rotational movement in being transferred from goal box to start box on test trials. Several hypotheses were offered in explanation of these results, and implications for the general conduct of latent learning experiments were suggested.

References

1. Bartlett, M. S. Square root transformation in analysis of variance. *J. Roy. statist. Soc. Suppl.*, 1936, **3**, 68–78.
2. Christie, R. Experimental naïveté and experiential naïveté. *Psychol. Bull.*, 1951, **48**, 327–339.
3. Christie, R. The role of drive discrimination in learning under irrelevant motivation. *J. exp. Psychol.*, 1951, **42**, 13–19.
4. Christie, R. The effect of some early experiences in the latent learning of adult rats. *J. exp. Psychol.*, 1952, **43**, 281–288.
5. Desiderato, O. An experimental investigation of latent learning. Unpublished master's thesis, New York Univer., 1951.
6. Honzik, C. H., & Tolman, E. C. The perception of spatial relations by the rat. *J. comp. Psychol.*, 1936, **22**, 287–318.
7. Jackson, T. A. General factors in transfer of training in the white rat. *Genet. psychol. Monogr.*, 1932, **11**, No. 1.
8. Karn, H. W., & Porter, J. M., Jr. The effects of certain pretraining procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461–469.
9. Ligon, E. M. A comparative study of certain incentives in the learning of the white rat. *Comp. psychol. Monogr.*, 1929, **2**, No. 6.
10. MacCorquodale, K., & Meehl, P. E. Preliminary suggestions as to a formalization of expectancy theory. *Psychol. Rev.*, 1953, **60**, 55–63.
11. Maltzman, I. M. An experimental study of learning under an irrelevant need. *J. exp. Psychol.*, 1950, **40**, 788–793.
12. Seward, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, **39**, 177–186.
13. Spence, K. W., Bergmann, G., & Lippitt, R. A study of simple learning under irrelevant motivational-reward conditions. *J. exp. Psychol.*, 1950, **40**, 539–551.
14. Thistlethwaite, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.
15. Vineberg, R. The relation of irrelevant-incentive learning to motivation and reward during exploratory experiences and training. *J. exp. Psychol.*, 1953, **46**, 237–242.

(Received September 6, 1955)

Journal of Experimental Psychology
Vol. 52, No. 4, 1956

EFFECT OF VARIOUS MOTIVATIONAL TECHNIQUES UPON LEARNING AND PERFORMANCE TASKS[1]

GEORGE H. ZIMNY

University of Minnesota[2]

The influence upon human activities of what is commonly called motivation is a very complex problem that is still in an exploratory stage. From an experimental point of view, the complexity of this problem is due to several factors. One factor is the many techniques of inducing motivation that can be and have been employed. Abel (1) used monetary and verbal reward, Hurlock (7) employed praise and reproof, Gilbert (6) used electric shock, and Alper (2) told *S*s that the task was a test of intelligence. Some *E*s have combined several specific techniques into one motivational variable. Russell (9) introduced to one group a combination of pride in one's school, comparison with other students, a reward of school credit, and information that performance was related to academic ability.

A second factor contributing to the complexity of the problem is the measured activities engaged in by *S*s. There are different types of activities and different measures of each type of activity. With regard to the former, Book and Norvell (4) used the task of writing *a*s and crossing out letters, Benton (3) gave intelligence tests, Russell (9) employed learning of nonsense syllables, and Travis (10) used a pursuitmeter. With regard to the latter, the various measures of performance such as response latency and number of trials to reach a criterion are well known.

A third factor that can be mentioned is the variety of specific conditions under which experiments on motivation are carried out. The number of such variations is enormous and this produces a definite problem since generalizations concerning these conditions are not now possible. In one study, Abel (1) used boys of above average intelligence in one part and boys of average intelligence in the other part. When nonsense syllables are to be learned on a memory drum, there are several specific conditions that can vary from experiment to experiment such as the speed of presentation, number of syllables, and meaningfulness of the syllables. The difficulties caused by this third factor form the basis for the often pleading request for standardization of experimental conditions in a particular area of investigation.

The present study deals with two problems arising from the factors mentioned above. The first part of the study (referred to as the learning experiment) is concerned with the effect of different motivational techniques introduced under similar experimental conditions. The general techniques used here have been used before in different investigations but each of these previous investigations has employed different specific conditions such as age of subjects or type of learning material. The purpose of

[1] This article is based on a dissertation submitted to the faculty of the Department of Psychology of the University of Minnesota in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Drs. W. A. Russell and M. A. Tinker for their aid in the planning and reporting of this study.

[2] Now at Marquette University.

the first part of this study is to determine the effect of these various motivational techniques when other conditions are kept constant.

The second part (referred to as the performance experiment) deals with the effect of a variation in type of activity upon the detectable effects of a specific technique of inducing motivation. It is possible that motivational effects due to a particular technique could be detected with one type of task and not with another. For example, it has been suggested (**5**) that motivation may exert differential effects upon verbal and motor activities. The present study employed both a learning task and a performance task (one involving little learning) under similar experimental and motivational conditions in order to investigate the possible influence of type of task.

Method

Subjects.—A total of 140 *Ss* used in both experiments were college men taking introductory laboratory psychology.

Apparatus and materials.—In the learning experiment, the material consisted of two lists of 12 three-letter nonsense syllables selected from lists prepared by Melton (**8**, Lists 7 and 8). The lists were presented serially on a memory drum and learning was by the anticipation method. Each syllable was exposed for 1.4 sec. followed by a blank space exposed for 1.4 sec. There was an interval of 7 sec. between each presentation of the list. A large screen blocked *S*'s view of *E*.

An apparatus capable of delivering an electric shock consisted of two dry-cell batteries, a rheostat, a telegraph key, and two wires all connected in series. The wires were to be attached to *S*'s fingers. This apparatus was covered on a table out of sight of *S*.

In the performance experiment, the apparatus consisted of 10 side-by-side compartments 3.5 in. wide and 4.5 in. long. Each compartment was marked with a randomly assigned number from 1 to 10. The numbers were covered when *Ss* were not actually sorting. Cards were thrown into the appropriate compartment and slid down out of *S*'s sight. Cardboard strips were available for covering the cards already sorted in order to get a measure of number of cards sorted per 2-min. interval. Two stop watches and an upright laboratory timer were used. Eleven decks of ordinary playing cards with the picture cards removed were the material to be sorted. Each deck contained 40 cards, four sets of from one (ace) to ten each. After being sorted, each deck was shuffled 10 times by *E*.

Procedure.—In the learning experiment, the 100 *Ss* were randomly assigned to four groups of 25 each. After being given the conventional instructions for learning nonsense syllables on the memory drum by the anticipation method, each *S* learned the first list to a criterion of six correct on one trial. The *Ss* were informed that this was a practice list. The instructions appropriate to each of the four groups were given verbally to each member of the group and the second list of 12 syllables was learned to a criterion of one errorless trial.

In the performance experiment, the 40 *Ss* were randomly assigned to two groups of 20 each. Following the initial instructions, each *S* stood in front of the sorting boxes and picked up the deck lying there. He turned up the cards one by one and put them into the appropriate boxes. The *Ss* were told that this first deck was a practice deck. The instructions for each of the two groups were given verbally to each member of the group. Ten decks of cards were laid out side by side in front of *S*. These decks were sorted one after the other until all 10 decks were completed.

Motivational instructions.—In the learning experiment, these instructions were introduced immediately before the second list was learned. The *Ss* in the Control Group were told that the next list would merely give them more practice in learning this kind of material. A 1-min. delay was introduced so that the time interval between the learning of the two lists would be the same for the Control Group as for the other three groups which received longer instructions.

For *Ss* in the Incentive Group the essential part of the instructions was as follows:

"Now on this list we want you to do the very best job you can, that is, learn the list as rapidly as possible. In order to make it worth your while to do this we have made arrangements with the psychology department to allow extra credit to those students who show their cooperation by learning this next list rapidly. You know you will receive credit for one 'r' report (long report) just for being in this experiment. Now, if you do well on the next list, it will be possible for you to receive credit not only for that but for still another 'r' report which can be handled just like the first. Thus, if you do well in learning

this list, it will be possible for you to be excused from two 'r' reports, one for being an *S* and one for learning this new list rapidly."

When each *S* in the Shock Group entered the experimental room, the table with the shock apparatus was moved out and placed next to *S*. Testing the apparatus produced an audible buzzing. The essential part of the instructions was as follows:

"The only difference is that on this list, every time I find that you are learning the list slower than you should be, you will receive an electric shock from this apparatus here. The electric shock is somewhat unpleasant but not harmful. I have used this list in previous studies and hence know approximately how fast it should be learned. Let me connect up the electrodes. [The two wires are attached to *S*'s fingers.] You should try to learn the list as rapidly as possible because the faster you learn it the surer you are of avoiding shock. Your goal should be to learn the list so rapidly that you are not shocked even once."

At no time was *S* shocked regardless of how slowly he learned the list. There was only threat of shock.

For *S*s in the IQ Group the essential part of the instructions was as follows:

"Several studies in the psychological literature have found that this new list which you are about to learn is correlated rather highly with one of the standard intelligence tests. The intelligence test used in these studies was the Wechsler-Bellevue, which is a test designed to determine the intellectual level of adults. In these studies the speed of learning scores for this list of nonsense syllables was obtained from a representative group of adults using essentially the same procedure as we are using here. Then the Wechsler-Bellevue intelligence test was administered to each person in the group and an intelligence test score determined. Correlating these two sets of scores produced a rather high, though not perfect, correlation. Thus, for all practical purposes, the speed with which you learn this new list will give an indication of your level of intelligence. In general, the faster you learn this list, the higher the level of intelligence it indicates that you have."

In the performance experiment, the motivational instructions were given immediately before the 10 decks of cards were sorted. The *S*s in the Control Group were told that the purpose of the additional decks was to give them more practice in card sorting. The *S*s were also asked several questions about school in order to make the instruction time equal to that of the Incentive Group.

The *S*s in the Incentive Group were given essentially the same instructions as those given to *S*s in the Incentive Group of the learning experiment.

Results

Learning Experiment

Practice list.—The *S*s in each of the four groups learned the practice list to a criterion of six syllables correctly anticipated on one trial before any experimental variation was introduced. Thus performance on the practice list provided a basis for comparing the four groups with respect to initial learning ability. The number of trials to reach the criterion was subjected to analysis by Bartlett's test for homogeneity of variance and by analysis of variance. In both cases the null hypothesis could not be rejected, thus indicating that the four groups did not differ significantly in initial learning ability on this task.

Experimental list.—All *S*s learned the experimental list of 12 syllables to a criterion of one perfect trial. Performance was measured in terms of the number of trials required to reach the criterion. Table 1 contains the means and *SD*s for the number of trials required to reach the criterion of 12 syllables correctly anticipated for the four groups. The Bartlett test for homogeneity of variance was carried out and the obtained chi-

TABLE 1

MEANS AND *SD*s OF THE NUMBER OF TRIALS TO REACH THE CRITERION OF 5 AND 12 CORRECT ANTICIPATIONS

(*N* in each group = 25)

Group	12 Correct		5 Correct	
	Mean	*SD*	Mean	*SD*
Control	26.32	11.47	7.12	2.89
Incentive	24.80	7.22	7.80	3.35
Threat of shock	27.88	10.98	8.76	4.56
Intelligence test	25.96	12.12	7.52	3.53

square value fell between the 5% and 10% levels of significance (approximately 8% level). The null hypothesis was not rejected. The *F* value obtained by the application of the analysis of variance was a fraction, thus indicating that the means of the four groups did not differ significantly. On the basis of these results, it was concluded that the motivational instructions produced no significant differential effects upon either mean performance or the variability of performance when the entire list was learned.

It is possible that motivational effects were present soon after the motivational instructions were given, but tended to "wear off" as *Ss* settled down to the grind of learning the entire list. To test the possibility of early motivational effects, an analysis was made of the number of trials required to reach a criterion of five syllables correctly anticipated (see Table 1). The Bartlett test for these values yielded a chi-square value lying between the 10% and 20% levels. Analysis of variance yielded an *F* value less than unity. It was concluded that the motivational instructions produced no significant differential effects at this early point in learning. Analysis of performance using a criterion of six syllables correctly anticipated led to the same conclusion.

The results obtained above are not in agreement with those obtained by Russell (9) using the same list of nonsense syllables and similar learning procedure. Russell found differences between a motivational and a control group in variability of performance but not in mean performance when learning proceeded to a criterion of five syllables correctly anticipated. It will be recalled, however, that Russell combined several motivational techniques into one motivational variable. Thus it is possible that a combination of techniques will produce significant variability in performance but that single techniques will not.

In order to permit a more complete comparison with the results obtained in the performance experiment, the data for the Control and the Incentive Groups were analyzed using a relative measure. This measure consisted of the difference between the number of trials required to reach the criterion on the practice list (six syllables correctly anticipated) and the number of trials required to reach the same criterion on the experimental list (zero and negative scores were obtained). For the Control Group the mean of the total number of trials difference was 6.56 trials ($SD = 6.26$). For the Incentive Group it was a mean of 4.36 trials ($SD = 4.83$). The *F* test for homogeneity of variance yielded a value below that required for significance at the 5% level. The *t* test for mean differences was applied and the value obtained was found to lie at approximately the 15% level. Thus the use of a relative measure led to the conclusion that the incentive instructions did not produce any significant differential effects upon either mean performance or the variability of performance.

Performance Experiment

Practice deck.—Initial card-sorting ability was determined by analyzing the number of seconds required by the Incentive and Control Groups to sort the practice deck. In this analysis, the *F* test for homogeneity of variance yielded a value below that required for significance at the 5% level. The *t* test for differences in means produced a *t* value lying between the 70% and 80% levels. In each case the null hypothesis was not rejected. Thus the two groups were considered equal in initial card sorting ability.

Experimental decks.—The *Ss* in each group sorted 10 decks of cards one

after the other following the instructions appropriate to their group. Performance on these 10 decks was measured in several ways.

Using two stop watches, a record was obtained of the number of seconds required to sort each deck. A comparison of the mean number of seconds sorting time required by each group for each deck revealed that the Incentive Group required less sorting time on every one of the 10 decks. The maximum difference of 3.7 sec. occurred on Deck 1. The next largest difference, 3.5 sec., occurred on Deck 7. Analysis by means of the F and t tests revealed no significant difference between the two groups in variability of performance or in mean performance for either Deck 1 or Deck 7.

Since the Incentive Group required fewer seconds sorting time on every deck, total sorting time was analyzed. For all 10 decks the Control Group had a mean sorting time of 712.20 sec. (SD = 72.88). The Incentive Group had a mean of 685.90 sec. (SD = 81.27). As above, analysis by means of the F and t tests failed to reveal any significant differences. Thus it was concluded that, in terms of this measure of performance, the incentive instructions had no significant effect upon initial or total performance in card sorting.

A measure was also obtained of the number of cards sorted per 2-min. interval. Once again, the largest difference between the two groups was in favor of the Incentive Group and occurred at the beginning of performance, namely, in the first 2-min. interval. This difference in mean number of cards sorted was 3.65 cards. The second largest difference in means, 3.60, was in the 6- to 8-min. interval. Analysis of each of these differences by the F and t tests produced no significant differences.

In agreement with what was found using the previous measure, the Incentive Group sorted more cards than the Control Group in each of the four 2-min. intervals for which complete data were available. In the 8-min. period the Control Group sorted a mean of 272.40 cards (SD = 30.42). The Incentive Group sorted a mean of 285.05 cards (SD = 37.99). Analysis by means of the F and t tests resulted in no significant differences. The t value fell at approximately the 25% level. The conclusion was drawn that the incentive instructions did not significantly affect initial or total card-sorting performance when this measure of performance was used.

The two measures of card-sorting performance used above were absolute measures and neither one yielded any significant indication of motivational effects. In searching further for possible motivational effects a third measure of performance was derived. This was a relative measure and consisted of the difference between the time required to sort the practice deck and the time required to sort each of the 10 experimental decks. Using this measure, comparison of the mean performance of each group on each of the 10 decks revealed the same pattern as found previously, namely, the Incentive Group performed better on every deck. The possibility is immediately presented that, in terms of the mean total decrease in time for all 10 experimental decks, the Incentive Group may have done significantly better than the Control Group. For the Control Group the mean of the total number of seconds difference between the practice deck and each of the 10 experimental decks was 122.30 sec. (SD = 53.05). For the Incentive Group it was a mean of 158.60 sec. (SD = 66.60). Application of the F and t tests resulted in an

F value lying below that required for significance at the 5% level and a *t* value lying between the 5% and 10% levels. The null hypothesis was not rejected in either case, indicating that when total performance was analyzed, the two groups did not differ significantly in either mean performance or variability of performance.

As was pointed out in connection with the learning experiment, it is possible that motivational effects were present early in the task but were not detected by the analysis of total performance. To test this possibility, a statistical analysis was made of the two groups' total performance on the first five decks using the relative measure described above. For the Control Group the mean of the total number of seconds difference between the practice deck and each of the first five experimental decks was 55.80 sec. (SD = 24.69). For the Incentive Group it was a mean of 75.05 sec. (SD = 31.22). Applying the *F* test for homogeneity of variance, the obtained *F* value was found to lie below the table value required for significance at the 5% level. The *t* test yielded a value that was significant between the 2% and 5% levels. Although this level of significance is such that caution must be exercised, it may still be concluded that when performance was determined by this relative measure, the motivational instructions tended to have a differential effect upon total mean performance on the first five decks of the card-sorting task, in a direction favoring the Incentive Group.

Since evidence of early motivational effects was found, it would be expected that the greatest motivational effect would occur immediately after presentation of the incentive instructions, that is, on Deck 1. Thus the mean difference scores for the two groups should show the greatest difference on Deck 1 and this difference should be significant. This expectation was borne out on both counts. The greatest difference among the 10 decks did occur on Deck 1. For the Control Group the mean of the total number of seconds difference between the practice deck and Deck 1 was 10.25 sec. (SD = 5.74). For the Incentive Group it was a mean of 14.95 sec. (SD = 6.53). Application of the *F* and *t* tests yielded an *F* value less than that required for significance at the 5% level and a *t* value significant at the 2% level. Thus the mean difference between the two groups may be considered significant, indicating that the incentive instructions were effective in facilitating performance on the first deck.

Discussion

The first problem investigated dealt with the effect of various types of motivational techniques introduced under similar experimental conditions. In the first part of the study, the results indicated that the motivational techniques employed did not have any detectable effect upon learning. There are several possible explanations for the absence of positive motivational effects: (*a*) the measures of performance used may not have been sensitive enough to detect motivational effects that were actually present; (*b*) the particular motivational techniques employed may have been too low on their own dimension of effectiveness; (*c*) it is possible that some or all of the *S*s did not sufficiently accept the motivational elements contained in the verbal instructions; and (*d*) the *S*s in the control group may have been so highly motivated by certain factors in the total experimental situation that whatever motivational effects were produced by the motivational techniques could not be detected when comparison was made to the control group. These possible factors further illustrate the complexity of the problem of the experimental investigation of human motivation.

The second problem investigated was the effect of type of task upon the detection of motivational effects. The effect of type of task was found to be dependent upon the measure of performance employed. When absolute measures were employed in both the learning and performance experiments, no motivational effects were detected, thus

indicating that type of task was not an important factor. However, when a relative measure was used in both experiments, initial motivational effects were detected in the performance experiment but not in the learning experiment, thus indicating that type of task was important. What specific difference or differences between the two tasks are important in this connection cannot be determined from this study. The tasks did differ in several ways. The learning task was a paced verbal task involving a great deal of learning whereas the performance task was an unpaced motor task involving little learning. Perhaps the verbal-motor difference is an important one as has already been suggested (5). The present study has shown, however, that the type of task can be important under certain conditions in the attempt to study motivational effects.

Summary

In the first part of the study, four groups of 25 college men each learned a list of 12 nonsense syllables after being subjected to one of the following conditions: (*a*) an incentive of an excused classroom assignment, (*b*) threat of electric shock, (*c*) statement that the task was a sign of intelligence, and (*d*) statement that the task afforded additional practice (control group). In the second part, two groups of 20 college men each sorted 10 decks of 40 cards each after one group had been subjected to the control condition and the other to the incentive condition described above.

The three techniques of inducing motivation had no demonstrable effects upon either mean performance or variability of performance in the paced verbal learning task when both absolute and relative measures were employed.

The incentive technique had no demonstrable effects upon either mean performance or variability of performance in the unpaced motor performance task when absolute measures of performance were employed. When a relative measure of performance was used, the incentive technique did not produce any detectable effect upon variability of performance in card sorting but it did produce a significant facilitating effect upon mean performance during the first half of the task.

These results of the learning and performance experiments indicated that the type of task used in the study of motivational effects can be an important variable, depending upon the measure of performance employed.

References

1. Abel, L. B. The effects of shift in motivation upon the learning of a sensorimotor task. *Arch. Psychol., N. Y.*, 1936 **29,** No. 205.
2. Alper, T. G. Task-orientation vs. ego-orientation in learning and retention. *Amer. J. Psychol.*, 1946, **59,** 236–248.
3. Benton, A. L. Influence of incentives upon intelligence test scores of school children. *J. genet. Psychol.*, 1936, **49,** 494–497.
4. Book, W. F., & Norvell, L. The will to learn—an experimental study of incentives in learning. *Ped. Sem.*, 1922, **29,** 305–362.
5. Farber, I. E., Russell, W. A., & Andreas, B. G. Effect of failure upon performance in verbal and motor learning situations. Paper read at Midwest. Psychol. Ass., Chicago, April, 1949.
6. Gilbert, R. W. The effect of non-informative shock upon maze learning and retention with human subjects. *J. exp. Psychol.*, 1936, **19,** 456–466.
7. Hurlock, E. B. The value of praise and reproof as incentives for children. *Arch. Psychol., N. Y.*, 1924, **11,** No. 71.
8. Melton, A. W. Materials for use in experimental studies of the learning and retention of verbal habits. Unpublished manuscript, Univer. of Missouri, 1940.
9. Russell, W. A. Retention of verbal material as a function of motivating instructions and experimentally-induced failure. *J. exp. Psychol.*, 1952, **43,** 207–216.
10. Travis, R. C. Comparison of the influence of monetary reward and electric shock on learning in eye-hand coordination. *J. exp. Psychol.*, 1938, **23,** 423–427.

(Received September 6, 1955)

Journal of Experimental Psychology
Vol. 52, No. 4, 1956

THE REINFORCEMENT DIFFERENCE LIMEN (RDL) FUNCTION FOR SHOCK REDUCTION[1]

BYRON A. CAMPBELL[2]

Harvard University

In a previous study (1) the reduction in white noise required to motivate "just noticeable" learning in rats was investigated as a function of the stimulus intensity reduced. These reductions, which may be called *reinforcement difference limens*, exhibited a relation to stimulus intensity quite comparable in form to the stimulus difference limen functions obtained in human psychophysical studies. The variance of the data from which these reinforcement difference limens (RDLs) were derived was small, and the relation between the size of the RDL and the intensity of the noise was essentially monotonic. The results suggest that the reinforcement difference limen may be of use in describing the motivating properties of different kinds of aversive stimuli, just as the stimulus difference limen has been of use in describing the relative discriminability of different stimuli in psychophysics. The purpose of the present experiment, accordingly, was to determine the relationship between the size of the RDL and stimulus intensity for another aversive stimulus, namely, electric shock.

METHOD

Subjects.—The *S*s were 150 rats of the Wistar strain. Their weights ranged from 300 to 550 gm.; 50 were females and 100 were males.

Apparatus.—The apparatus consisted of a simple tilting cage 18 in. long, 8 in. wide, and 8 in. high. When *S* was in one half of the cage, a standard level shock intensity was turned on and maintained by a tilt-operated microswitch. When *S* moved to the opposite half, the cage tilted and the opening or closing of the microswitch turned on a lower (comparison-shock) intensity which remained on as long as *S* stayed in that half of the cage. If the difference between the standard shock intensity and the comparison shock intensity was large, *S* readily learned to stay in the lower-shock half of the cage; if the difference was small, essentially chance preference for the two halves resulted. The time *S* spent per minute in the lower-shock half of the cage and number of crossings from one half to the other were recorded on a polygraph.

The shock circuit consisted of two radio transformers, each with a 20,000-ohm continuously variable potentiometer across the output. One supplied the standard shock voltages and the other the comparison shock voltages. In series with the rat through a common ground lead was a 150,000-ohm resistor. This resistor limited the total current flow through the rat and made the current flow relatively constant. In calculating the intensity of shock stimulation reaching the rat, it was assumed that the effective stimulus intensity from shock is proportional to the power dissipated across the rat where $P = E^2/R$. Since it is extremely difficult to measure the actual voltage dropped across the rat, it was assumed that this voltage would be directly proportional to the voltage read between the potentiometer arm and ground even though a considerable portion of the voltage was actually dropped across the resistor in series with the rat.

This resistor merely cuts down the voltage imposed on the rat by a certain fraction and does not change the basic relationship. As a corollary of this, it was assumed that the average resistances of the rats in the different groups were equal,[3] and as a constant did not need to be

[1] This research was conducted while the author was a National Science Foundation Postdoctoral Fellow at Harvard University, and he wishes to thank the National Science Foundation and the Psychology Department of Harvard University for supporting this research.

[2] Now at Princeton University.

[3] Research by the author now in progress indicates that this assumption is not correct. Instead it appears that the resistance of the rat decreases as shock intensity increases. Accordingly, the findings reported in this article can only be considered an approximation of the

considered in the equation $P = E^2/R$. Variations in the rats' resistance would add only to the error variance in this formulation, and the functional relationship between the estimated electrical power dissipated in the rat and his behavior would be unchanged. Thus it was assumed that the square of the voltage measured at the output of the voltage divider before the series resistor was directly proportional to the power dissipated in the rat. This measure of shock intensity was used throughout the study.

Design and procedure.—The over-all plan was to determine for a number of different shock intensities the amount of shock reduction required from those intensities required to produce "just noticeable" learning. "Just noticeable" learning was defined as the acquisition of a 75% preference for the lower stimulus-intensity half of the tilting cage. In all, there were six "standard" shock levels of 55, 72, 95, 165, 287 and 500 v., respectively. These standard stimuli were equidistant on a log scale of E^2 (except for the second standard which was halfway between the first and third standard intensities). For each standard level there were at least three comparison levels of a lower shock intensity; the comparison stimuli for each standard were also equidistant on a log scale of intensity. The comparison levels were selected on the basis of preliminary research, and the intent was to bracket the 75% preference level. The preference data for these groups could then be plotted as a function of the standard shock intensity, and the intensity reduction producing a 75% preference could be determined from this plot. This method is similar to the psychophysicists' "method of constant stimuli."

Six *S*s were run for each of the different comparison stimuli of a given standard intensity, and each *S* was run only once to avoid "sequence" effects. (Preliminary research indicated that responses learned in the first shock combination often perseverated during attempts to train *S* with new shock combinations.) Each *S*'s session in the tilting cage lasted for 8 min. Any preference for a particular half of the cage was balanced out by running half of the *S*s in each group with the standard (higher) intensity at one end and half with the standard intensity at the other end of the tilting cage. Animals were assigned randomly to the various experimental conditions.

The *E*'s procedure in running *S* was to take *S* from its group living cage, place it in the tilting cage with no shock present, and go to the adjoining room, where the shock-producing and recording apparatus were housed. The *E* then set the shock output to the appropriate intensities and closed the switch activating the grid floor of the cage. Then *E* returned to a position where he could observe *S* and remove any feces which shorted the grid. At the end of the session *S* was removed from the tilting cage, and the stainless-steel grid floor was wiped clean with a damp paper towel.

relationship between the magnitude of the RDL and shock intensity.

Results

As a first step the data for a given standard stimulus were analyzed by plotting the percentage of time spent at the lower-shock half of the tilting cage against the square of the voltage of the lower-shock half. This analysis for the six standard stimuli is shown in Fig. 1. The measure used at each comparison point in the figure was the median percentage of the time spent in the lower-shock half during the last four minutes of the 8-min. trial. The solid black circles along the abscissa at the 50% preference level indicate the shock intensities at which the standard level intensities equal the comparison level intensities. (It is assumed that with a sufficient

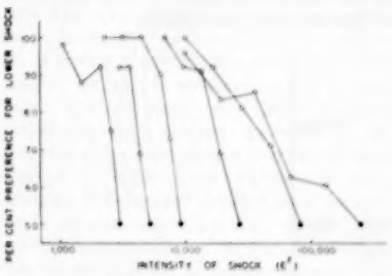

Fig. 1. Per cent preference for the comparison shock (lower shock half of the tilting cage) plotted as a function of comparison shock intensity. The open circles depict the comparison shock intensities and the associated preferences for each of those intensities. The filled circles indicate the shock intensities at which the comparison level intensities equal the standard level intensities. It is assumed that a preference of 50% for the comparison shock will result when the comparison and standard intensities are equal.

N, a 50% preference for the comparison shock side would result when the test and standard shock intensities were equal.) The hollow circles indicate the obtained preferences for the comparison shocks. The greater the difference between the standard level intensities (black circles) and the comparison level intensities (open circles) the greater the preference for the comparison shock.

The next step in the analysis of these data was to determine RDLs for each standard level, i.e., the reduction which was "just noticeably reinforcing" in that it produced a 75% preference for the lower shock. This was done by fitting curves to the data shown in Fig. 1, using the "normal graphic" method (**3**), and by determining the point for each standard level, where the function crossed the 75% preference level. The RDLs determined in this manner are plotted as a function of log stimulus intensity in Fig. 2. The plotted relation indicates clearly that the RDL (in log units) increases linearly as a function of log standard intensity.

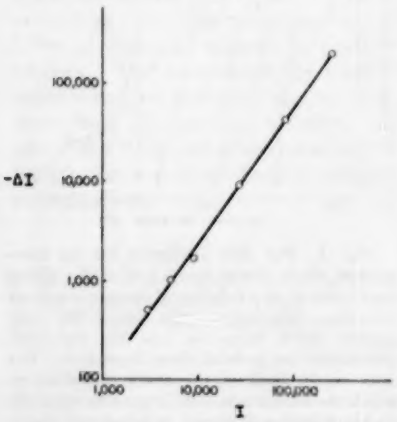

FIG. 2. Reinforcement difference limens ($-\Delta I$) plotted as a function of log stimulus intensity. Both axes are in E^2 units.

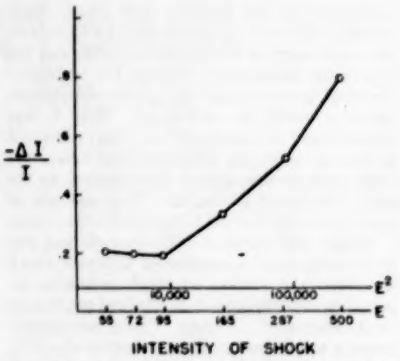

FIG. 3. Relative RDLs or "reinforcement fractions" plotted as a function of shock intensity.

The final step in the analysis of the results was to plot the relative RDL function—the counterpart to the psychophysicists' relative DL or Weber function. This was done simply by dividing the absolute intensity reduction ($-\Delta I$) required to produce a 75% preference by the intensity (I) of the standard level. These fractions ($-\Delta I/I$) are plotted as a function of stimulus intensity in Fig. 3 to achieve the relative RDL function for shock as a motivator of learning. The above plots are completely analogous to the DL functions of psychophysics: the criterion of a "just noticeable" difference for the RDL is whether the difference (reduction) is reinforcing, whereas for the DL the criterion is whether the difference can be detected.

Another set of data which are of interest show the acquisition of the preference response. Figure 4 depicts the acquisition curves of the fourth standard level ($E = 165$). This level was selected as a typical standard level in lieu of attempting to present acquisition curves for all of the 25 groups employed in the study. The data show that acquisition is rapid for the maximal reductions and is nearly

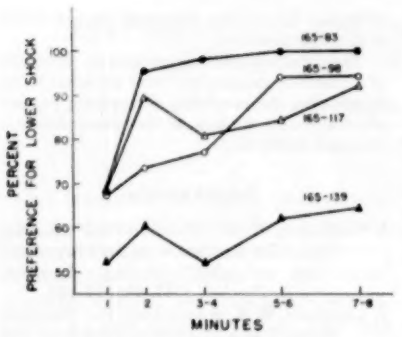

FIG. 4. Acquisition curves for the 165-v. standard level. The numbers adjacent to each curve indicate the voltages of the standard and comparison levels, respectively.

asymptotic for all groups during the last 4 min. of the training period. Acquisition curves for the groups run at the remaining standard levels are much the same.

DISCUSSION

The present results agree with the previous study (1) using white noise in the finding that stronger aversive stimuli required larger *absolute* amounts of intensity reduction in order to produce reinforcement. This general fact is best shown for the present study in Fig. 2, which displays a linear relation between $-\Delta I$ and I for intensity reductions producing just noticeable learning. This result is also consistent with an earlier shock study (2) demonstrating that a constant decrement in shock intensity had a smaller reinforcing effect the larger the starting intensity level. The present study can be regarded as an elaboration and quantification of this general finding in the earlier shock study.

The relative RDL function shows that not only does the absolute decrement required for learning increase with greater shock intensities but also that the proportional decrement required for learning increases as a monotonic, positively accelerating function of intensity level. This function, shown in Fig. 3, is fairly flat at low intensities where the ratio is about .2, but it exhibits a steep rise to about .8 over the range of intensities (100 to 400 v.) generally used in instrumental learning studies utilizing shock reduction as an incentive.

This result from the present study with shock is in striking contrast with the relative RDL function in the previous comparable experiment using white noise as an aversive stimulus to motivate learning (1). In the latter study the relative RDL started at a high ratio of about .9 for the weakest effective sound levels and fell slightly over the available range to .7 at the highest levels used (115 db.). If the results of the two studies are taken to represent a continuum of aversive stimulation the implication is that the over-all curve relating stimulus intensity to the size of the relative RDL first drops off and then rises with the noise RDLs representing the dropping portion and the shock RDLs representing the rising portion. This shape function is one commonly found in psychophysical studies when relative DLs are plotted as a function of stimulus intensity.

The question will naturally arise as to whether the present results provide a new basis for scaling shock intensities in learning experiments utilizing shock escape or shock avoidance. It is possible, for example, to apply classical psychophysical scaling methods to the present data and obtain a scale of summated j.n.r.d.s (just noticeably reinforcing differences) similar to Fechner's scale of summated j.n.d.s. If this procedure is carried out on the above data on shock a scale of reinforcement consisting of approximately 15 j.n.r.d.s is obtained over the range from 55 to 500 v. Using this scale it is possible to describe the reinforcing properties of a given shock in terms of the number of j.n.r.d.s it is above the reinforcement threshold. Whether or not this proves to be a useful scale of shock incentive depends upon the extent to which it can be shown to integrate shock incentive studies using other performance indices

than the tilting cage preference used in the present study.

Summary

The present experiment made use of psychophysical methods and concepts to study the reinforcing properties of reductions in shock-produced pain. A research design similar to the psychophysicist's method of constant stimuli was used to determine the relationship between the size of the reinforcement difference limen (the reduction in stimulus intensity required to produce a 75% preference at the end of training for the lower shock half of a tilting cage) and stimulus intensity.

The results of the present experiment indicated that the size of the RDL (in log units) increased linearly as a function of log stimulus intensity. Further analysis of the data showed that the size of the relative reinforcement difference limen also increased as a function of stimulus intensity.

These findings were compared with the results of a previous comparable study on white noise. In addition, the possibility of applying psychophysical scaling methods to the present data was discussed briefly.

References

1. Campbell, B. A. The fractional reduction in noxious stimulation required to produce "just noticeable" learning. *J. comp. physiol. Psychol.*, 1955, **48**, 141–148.
2. Campbell, B. A., & Kraeling, D. Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, 1953, **45**, 97–101.
3. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.

(Received September 16, 1955)

Journal of Experimental Psychology
Vol. 52, No. 4, 1956

PSYCHOLOGICAL FACTORS IN MUSCLE-ACTION POTENTIALS: EMG GRADIENTS [1]

WALTER W. SURWILLO

Allan Memorial Institute and McGill University

An analysis of much current research shows that it suffers generally from a failure to view motor behavior in systematic relation to the traditionally defined psychological processes. At the level of theory this has been reflected in a trend toward either rejection of concepts like motivation and emotion, or consolidation of them into some broad, general category (**8, 9**). In the muscle-action-potential studies of motor behavior, on the other hand, we find frequent reference to and use of the concepts of motivation, attention, emotion, and effort (**11, 14, 15, 19, 21, 25**). However, while there is much evidence which points to the existence of a relation between electrical activity recorded from the skeletal musculature and such processes, little in the way of an explicit formulation has emerged. This study was designed to investigate the nature and extent of some of these relations.

Paralleling routine use in this laboratory of ink-writing electromyographs and integrators, interest has focused on the time-course variations of muscle-action potentials (EMGs) recorded during motor performance. Studies of muscular tension during mirror-tracing have reported progressive and systematic increases in the EMGs recorded from muscles active in the task and from inactive or passive muscles (**1, 2, 18, 23**). Referred to as the *EMG gradient*, this progressive increase in frequency and amplitude of muscle action potentials was found to begin with initiation of the task and to continue to its completion.

While the factors related to their occurrence have not been isolated, available data suggest that EMG gradients may be related to several important psychological variables. On the basis of his finding that *S*s who drew around the circle faster or performed the mirror tracing more accurately showed steeper gradients, Bartoshuk (**1**) postulated that slope of the EMG gradient reflects and varies with the degree of "effort" expended in performing the task. In later publication, these differences in slope were inferred to reflect differences in *S*'s level of "motivation" (**3**). Smith (**23**), on the other hand, suggested that the goal—or some discrete terminal point in the task—was the factor determining this increase. By implication, gradients should not appear or should be less marked in tasks where a goal is not apparent to *S*. Difficulty of a task has also been associated with increases in muscle-action potentials (**6, 7, 22**). The "build-up" phenomenon observed by Davis (**6**) seems related to the EMG gradient although it is not clear if

[1] This work was supported by the Defence Research Board of Canada, Grant Number 9425-04, Project D50-94-25-04, and by the Research and Development Division, Office of the Surgeon General, Department of the U. S. Army, under Contract No. DA-49-007-MD-626. This paper is based on a thesis submitted to the Faculty of Graduate Studies and Research, McGill University, in partial fulfillment of the requirements for the Ph.D. degree. Acknowledgment is made of the generous advice and support given by Dr. Robert B. Malmo.

both instances of increased activity are one and the same phenomenon.

The purpose of the study reported here was to test directly for a relationship between EMG gradients and level of motivation, difficulty, and goal structuring of a task by manipulating these factors experimentally. A motor task designed to limit muscular response to nearly isometric contractions and to control amount of contraction was studied. Two tasks—Task A and Task B—were administered to a group of *S*s. While both required identical muscular responses, incentives were considerably higher in Task A, it was more difficult to perform, and more strongly goal structured than Task B. In a second experiment an attempt was made to isolate the most relevant of these factors.

Experiment I

Method

Apparatus.—A tracking task in which *S* positioned a pointer by applying different amounts of rotational force to a controller knob was used. The display unit of this apparatus provided for either pursuit or compensatory tracking and consisted of two identical 4-in. microammeters mounted back to back and arranged so that the pointers fell in a horizontal line when each rested at the mid-point of its scale. The controller knob was positioned below the two meters where a seated *S* could grasp it comfortably in the same fashion as a doorknob. Rotation in the counterclockwise direction (pronation of the wrist and forearm) produced a corresponding movement of the right-hand pointer from its normal resting position near the top of the scale. Pointer deflection and angular rotation of the knob were proportionally related; these in turn were directly proportional to the applied force since rotation was opposed by a restraining spring. Maximum deflection of the pointer required a force of approximately 1200 gm. and a 10° rotation of the knob. This corresponded to nearly isometric conditions of muscular contraction. Systematic increases in electrical activity, therefore, could not be attributed to the physical characteristics of the task since under isometric conditions the electrical activity recorded from human muscle parallels and is directly proportional to mechanically recorded tension (**13, 17**).

The position of the left-hand pointer was not under the control of *S* but could be varied remotely by means of a function generator. During pursuit tracking, this pointer moved from its rest position down the scale at a constant rate to a point of maximum deflection and then back again. This movement was repeated cyclically with each cycle lasting for 30 sec. One complete trial of pursuit tracking consisted of five such cycles and *S*'s task was to make the position of the right-hand pointer coincide at all times with the position of the one on the left. With compensatory tracking the same cycles were applied instead to the right-hand pointer but in the opposite direction. In this case *S* was required to hold the pointer within a narrow sector marked on the scale. Both tasks required equivalent muscular responses.

Electrically, the tracking system consisted of two identical Wheatstone-bridge circuits. The output of each bridge was connected, respectively, to one of the display meters. In the case of compensatory tracking, two linear potentiometers made up the adjacent arms of one bridge circuit. The shaft of one potentiometer was connected mechanically to the shaft of the controller knob and rotated with it, while the second formed part of the function generator which was simply a motor arranged to drive a cam at the rate of 2 rpm. The cam rotated the shaft of this potentiometer at a constant rate between two set limits. With pursuit tracking, each potentiometer formed one arm of a different bridge circuit. A measure of the accuracy of performance was obtained by amplifying and recording the voltage which appeared across the controller-knob potentiometer.

The controller knob also included a device which was sensitive to variations in *S*'s grip pressure. While testing the apparatus, it was found that electrical activity recorded from electrodes placed on the ventral surface of the wrist (over *pronator quadratus*) and over the biceps muscle was some function of the force with which *S* gripped the controller knob. Grip pressure, therefore, was recorded as a control measure to insure that EMG gradients were not merely the result of systematic increases in grip during performance. The design and operation of the grip-pressure-sensitive system has been described elsewhere in detail (**24**).

Muscle-action potential recording.—Muscle-action potentials were recorded on an Offner Type D Electroencephalograph. Closely spaced bipolar surface electrodes were used for all recordings. Details regarding the electrodes and

placement techniques used in this laboratory have been reported previously by Davis (**5**).

Activity from all the muscle groups studied was rectified (full-wave rectification) and integrated electronically by means of circuits operating on the condenser-charge-discharge principle. These circuits were connected in parallel with the Offner galvanometers. Integrations were at the rate of 2 per sec., i.e., the condensers were periodically allowed to charge and then to discharge for periods of .48 and .02 sec., respectively. The magnitude of charge on the condenser for any .5-sec. period was recorded as a monophasic spike on a chart. Amplitude of each spike thus represented average electrical activity over a .48-sec. interval.

Since the tracking task engaged mainly the pronator muscles and their antagonist, the biceps, electromyograms were recorded from the following areas: ventral surface of the right wrist (*pronator quadratus*), right biceps (*biceps brachii*), and ventral surface of the left wrist (*pronator quadratus*). EMGs from the forehead (frontal belly of the *occipitofrontalis* muscle) were also observed since EMG gradients have been reported from this recording site (**26**). The reference or grounded electrode was placed on the dorsal surface of the left wrist.

Test structure.—The factors of incentive, difficulty, and goal structuring were studied by applying them to two tasks in such a fashion that Task A was weighted "high" with respect to these factors and Task B was weighted "low."

Task A was presented to *S* as *the* task, and he was informed that there would be money rewards for accurate tracking. In contrast, Task B was unrewarded and was not administered as a task: *E* merely told *S* in an informal manner that he was helping with some routine checks of the apparatus for the test to follow. Rewards were scaled roughly to permit *S*s to earn up to $1.00 in addition to the flat sum paid for serving in the experiment. More money could be earned (in increments of 20 cents) by keeping the right-hand pointer within the narrow sector marked on the scale at all times. In practice, the largest reward paid to any *S* was 80 cents. Quality of performance was judged roughly by visual inspection of the performance record traced on the chart.

Difficulty was scaled in a similar "high-low" fashion by administering compensatory tracking as Task A and pursuit tracking as Task B. Available data indicate that, other conditions remaining constant, better performance scores are achieved in pursuit than in compensatory tracking (**10**).

To accentuate the goal structuring in Task A, *S* was informed at regular intervals regarding the amount of the task completed and the amount remaining. This was accomplished by successively illuminating and extinguishing, from left to right, one of a series of colored lamps which was mounted in a row above the two meters. As the trial progressed, the position of the single illuminated lamp shifted periodically to the right. The end of the trial was signaled when the last lamp in the series was lighted. Task B, on the other hand, was performed merely with the instructions that it would last from 2.0 to 2.5 min.

A third task was administered to observe the effects of muscular fatigue on the action-potential responses. Some control for muscular fatigue seemed necessary in this study since the available data indicate that, under conditions where muscular effort is held constant, action-potential amplitude is greater at fatigue than at the time of initial contraction (**4, 16, 20**). These findings, of course, raised the possibility that EMG gradients were the result of muscular fatigue. Therefore in this third task—the Fatigue Control—*S* turned the controller knob to produce a maximum deflection of the right-hand pointer, and then held the pointer in this position for 2.5 min. (the length of the other tasks).

Testing procedure.—After rapport with *S* was established, electrodes were applied and routine checks were made of the recording system. Instructions were given regarding *S*'s position during testing; care was taken to insure that all *S*s maintained the same posture. This served to reduce the variability of the resting EMGs. The *S*s were given several opportunities to manipulate the controller knob and to observe its operation before the testing began.

Each trial of every task started with a 30-sec. interval of relaxed attention in which *S* assumed the exact position maintained throughout the trial. A lamp located at the left side of the display signaled this preperiod, and EMG recording was simultaneously initiated. The 2.5-min. period of performance was followed by a 30-sec. interval of relaxed attention corresponding to the preperiod. Extinguishing all colored lamps signaled the end of the postperiod and the beginning of an interval of freedom of movement for *S*. Rest periods of approximately 5 min. separated such trials.

Two trials of both Task A and Task B were given in balanced order to control for practice effects. After the four trials were completed, a single trial of the Fatigue Control was administered. When Task B followed Task A, *S* was told that the "routine checks of the apparatus" in which he was allegedly assisting were for the task to follow, i.e., the Fatigue Control. The latter task therefore served a dual function.

Subjects.—Sixteen right-handed male college

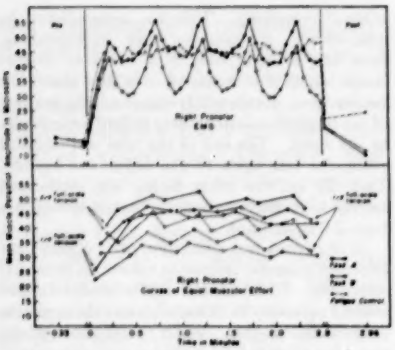

FIG. 1. The upper graph shows EMGs recorded from *pronator quadratus* (right wrist). The lower curves are derived from those shown above by connecting points where equal force is applied to the controller knob by *S*. Full-scale tension refers to the maximum force (1200 gm.) required at the peak of each cycle, and corresponds to maximum (10 degree) counterclockwise rotation of the knob. $N = 15$.

students (undergraduates) were tested. Seven were administered Task A first, nine Task B.

General treatment of data.—All measurements of muscle-action potentials were made from the integrator records. The primary EMGs from the Offner were inspected routinely and were used to check for and eliminate artifacts such as 60-cycle "hum," electrocardiograms, movements, and GSRs from the integrator records. Integrator spikes were measured in millimeters; these numbers were subsequently converted into microvolt values by reference to calibration standards. Results from the two trials of each task were averaged.

Two integrator spikes were measured from the 30-sec. preperiod. The first was taken approximately 10 sec. after the start and the second approximately 2 sec. before the end of the preperiod.

The 2.5 min. for which the tasks ran were scaled off into 41 intervals so that intervals 1–9, 9–17, 17–25, 25–33, and 33–41 each corresponded to one of the tasks' five cycles. Single integrator spikes were measured at each of the 41 points. Spikes measured at 1, 9, 17, 25, 33, and 41 were points where minimum tension was applied to the controller knob, while those measured at 5, 13, 21, 29, and 37 corresponded to maximum tension. Similarly, spikes lying midway between 1 and 5 (i.e., 3), 5 and 9, 9 and 13, etc., represented one-half maximum (full-scale) tension. In this fashion, curves of muscle action potential were obtained with muscular effort held constant.

Two integrator spikes were measured from the 30-sec. postperiod. The first occurred 2 sec. after the task had been completed and the second 2 sec. before the end of the postperiod.

Results

The upper portions of Fig. 1 and 2 show the average EMGs from the active muscles for the group studied. Both curves describe a cyclic rise and fall in activity which patterned the time course of the tracking function. The gradients shown in the lower portions of these figures were not linear with respect to time but appeared to follow some exponential function. Considering only the total rise from beginning to end of performance, the gradients from both right pronator and biceps are statistically reliable by Wilcoxon's (**27**) nonparametric test for paired replicates (*P*s range between .01 and .05).

Right-pronator gradients for Task A are all significantly steeper ($P < .01$) than those observed in Task B. Those for Task A are, in turn, steeper ($P < .05$) than the action-potential rise shown for the Fatigue Control. In the case of the biceps muscle, these

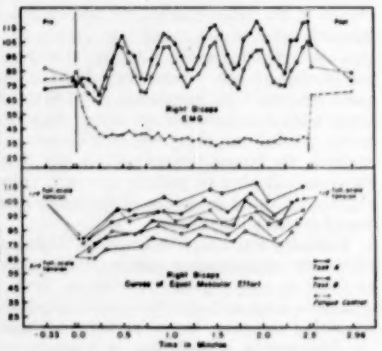

FIG. 2. The upper graph shows EMGs recorded from *biceps brachii* (right biceps muscle). Curves in the lower portion of the figure show the EMG gradients and are derived in the same fashion as those in Fig. 1. $N = 16$.

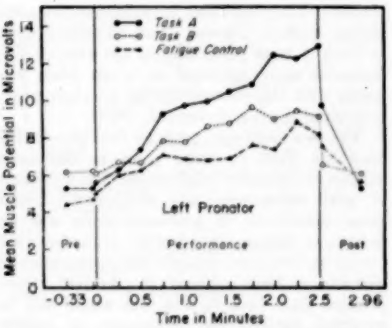

FIG. 3. EMG gradients recorded from *pronator quadratus* (left, passive wrist). Task A, compensatory tracking; Task B, pursuit tracking. $N = 15$.

differences in slope barely reach the .05 level of significance.

Results from the passive muscles are shown in Fig. 3 and 4. Left-pronator gradients were linear with respect to time (chi-square test shows that deviation from linearity is not significant) and highly significant (Ps range between .01 and .02). The rise for Task A is steeper than that shown for Task B and Fatigue Control ($P < .01$). Forehead data showed statistically reliable increases for Task A and Fatigue Control. Differences in slope for the three conditions were not significant.

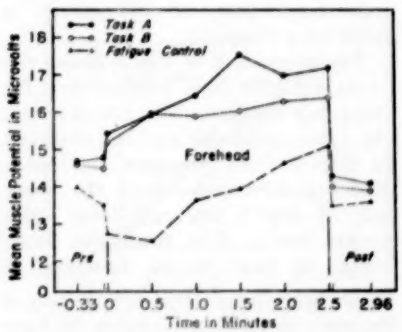

FIG. 4. EMG gradients recorded from frontal belly of the *occipitofrontalis* muscle (forehead). $N = 16$.

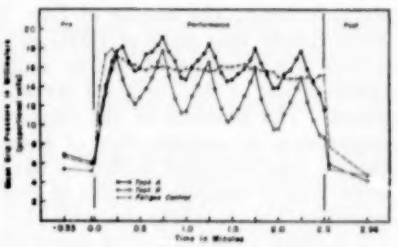

FIG. 5. Strain-gage recordings of grip pressure. Note the *negative* gradients. $N = 15$.

Figure 5 summarizes the grip-pressure data and shows that gradients do not occur in this measure: initial and final points corresponding to $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ full-scale tension do not differ statistically. Instead, reliable *negative* gradients appeared. The continuous fall in grip pressure from the second to the fifth peak, and from the second to the sixth trough is significant for both Tasks A and B ($P = .02$ and .01, respectively).

Performance scores were determined by comparing the records obtained with "perfect" performance and integrating the deviatiosn throughout the course of each trial with a planimeter. Expressed in units of square centimeters, the mean score ($N = 16$) for Task A was 41.3 sq. cm.; for Task B, 19.8 sq. cm. This difference is very significant ($P < .001$).

EXPERIMENT 2

While previously observed EMG gradients were reported as linear functions of time (**1, 2, 3, 23**), those recorded in Exp. 1 were nonlinear, with the exception of the results from the passive muscles. As the effort requirements of performance in this study were probably greater than in mirror tracing, it was desirable to repeat the experiment under conditions where this factor was reduced.

With the exception of the forehead

data, however, the results of Exp. 1 show clearly that slope of EMG gradients is a function of one or some combination of the variables investigated. The problem of the relative importance or contribution of these factors was attacked with the data from Exp. 1 through the use of correlations. Since it was probable that *S*'s response to the incentive would be reflected in his performance, correlations between slope of EMG gradients and performance scores were computed. None of these correlations proved to be significant. As the validity of this technique hinged on the assumption that the task was equally difficult for all *S*s, the question was attacked directly by investigating the effect of different levels of incentive.

Method

Apparatus.—The apparatus used was identical to that in Exp. 1 except that (*a*) the applied force corresponding to a maximum deflection of the right-hand pointer was reduced from 1200 gm. to 300 gm. (10° rotation unchanged), and (*b*) a device was included which furnished performance scores for compensatory tracking in terms of the more conventional "time-off-target" units. This consisted of a Contact Meter connected (via a buffer amplifier) in parallel with *S*'s meter and adjusted to close an internal switch whenever the pointer on *S*'s meter fell outside the target area. The switch activated an electronic contact relay similar to that described by Hunter (12), which in turn operated a chronoscope. With this device an accurate performance score could be obtained immediately after *S* completed the task.

Test structure and procedure.—Two trials of both Task A and B and one trial of the Fatigue Control were administered in the manner already described. After these were completed, *E* told *S* that he was interested in learning "how well you can really do on the test." Correspondingly, he was informed of his average score on Task A in terms of seconds "off target" and offered the opportunity to earn, in addition to his reward, a sum equal to twice that value by spending an additional 10 or 15 min. in testing. Thus, if his reward amounted to 80 cents, he could earn $1.60 extra (all *S*s agreed to the terms). The conditions for receiving the extra money were as follows: Two additional trials of Task A were performed; if the average performance score achieved on these trials was better than the average for the previous trials, *S* received the double reward.

The two additional trials of Task A—doubly rewarded Task A—were given in the same manner as the earlier trials except that the factor of goal structuring was eliminated. Under these conditions, if goal structuring was an important factor, the doubly rewarded task would be weighted against the appearance of steeper gradients. Since it received more practice than Task A, this task would also be weighted against the appearance of steeper gradients if difficulty proved to be an important factor. Compared with Task A, therefore, incentives were higher in doubly rewarded Task A, it was less difficult, and less strongly goal structured.

Muscle-action-potential recording.—The same EMGs observed in Exp. 1 were recorded. A fifth muscle, *pronator teres* (ventral surface of the right forearm), was included for study.

Subjects.—Sixteen right-handed RCAF men were tested.[3] All were ground-crew personnel with a mean age of 25 yr. Half of the group was administered Task A first, the other half Task B.

Results

Active muscles.—Curves from right pronator and biceps are not shown since they are similar to those recorded in Fig. 1 and 2. Gradients from both of these muscles were exponential functions of time. While the gradients for the three conditions were significant, no differences in slope were observed.

The upper half of Fig. 6 shows the average EMGs from *pronator teres* for the group studied. The curves reveal the same cyclic rise and fall observed in EMGs from *pronator quadratus*. EMG gradients shown in the lower half of Fig. 6 are not linear (chi-square test). The total rise from initial to final points, however, is

[3] We wish to acknowledge the cooperation of the men from the Royal Canadian Air Force who served as *S*s and to thank the officers who made the arrangements possible.

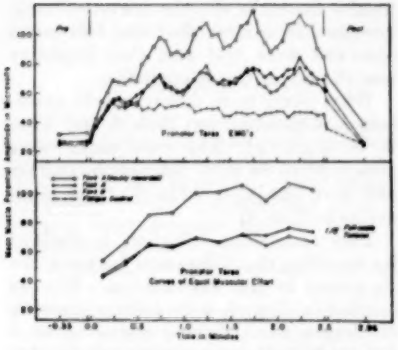

FIG. 6. The upper graph shows the EMGs recorded from *pronator teres* (right forearm). Curves in the lower portion of the figure are derived in the same fashion as those in Fig. 1. Full-scale tension refers to the maximum force (300 gm.) required at the peak of each cycle, and corresponds to maximum (10 degree) rotation of the knob. $N = 16$.

significant beyond the .01 level. Slopes for Task A and Task B do not differ significantly; slope for doubly rewarded Task A, however, is steeper than the slopes in either of the two conditions ($P < .01$). Gradients for $\frac{1}{4}$ and $\frac{3}{4}$ full-scale tension are not shown but follow the same pattern.

Passive muscles.—Results from the left pronator for the group are shown in Fig. 7. With the exception of the curve for Task B, all were linear functions of time. EMG gradients for the three tasks are very significant ($P < .01$); the rise for Fatigue Control is minimally significant ($P = .05$).

The slopes of the curves for Task A and B do not differ significantly; however, the gradient for doubly rewarded Task A is steeper than either curve ($P = .05$). All three curves are realiably steeper than that for Fatigue Control ($P < .01$).

Curves for the forehead are not shown since the only significant rise observed from this recording site was for the Fatigue Control.

Grip pressure and performance data.—Figure 8 summarizes the measures of grip pressure. As in Exp. 1, negative gradients again were apparent; in this case, however, they were not statistically reliable.

Mean performance score for Task A was 18.99 sec. "off target," and for doubly rewarded Task A, 9.03 sec. These means differ significantly at the .01 level.

The two trials of doubly rewarded Task A were split into two groups according to performance scores. Trials which yielded the "best" performance were collected into one group while the others—which, by comparison, showed the "worst" performance—formed the other group. Six *Ss* did "best" on the first trial

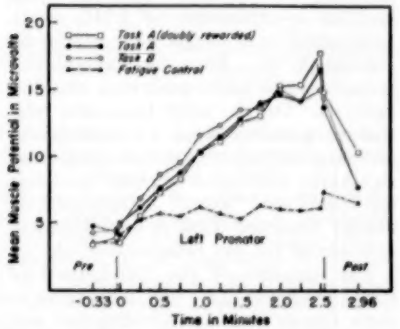

FIG. 7. EMG gradients recorded from *pronator quadratus* (left, passive wrist). $N = 15$.

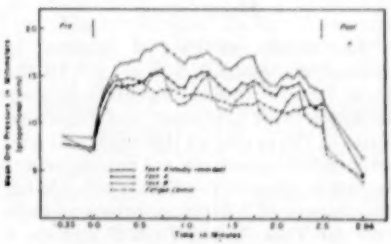

FIG. 8. Strain-gage recordings of grip pressure. $N = 16$.

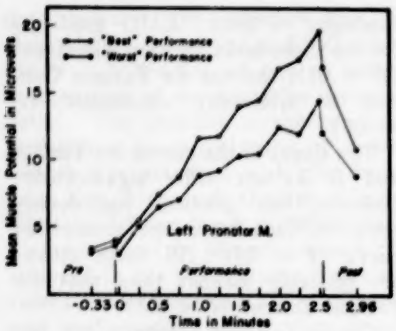

FIG. 9. EMG gradients from *pronator quadratus* (left wrist) for "best" and "worst" performance on doubly rewarded Task A. Average scores for best performance, 4.57 sec. "off target"; for worst performance, 13.49 sec. "off target." $N = 16$.

while the remaining 10 performed best on the second. The EMGs were grouped in this manner to determine whether steeper EMG gradients appeared with better performance.

Figure 9 shows the curves obtained from the left pronator data. The EMG gradients are linear; the one recorded for "best" performance is significantly steeper ($P = .05$) than the corresponding gradient for "worst" performance.

Data from the four other muscles and the grip-pressure measures were treated in the same manner. No reliable differences were found between the two conditions.

DISCUSSION

The results of Exp. 2 confirm, in general, the findings in Exp. 1. Of the three factors studied, incentive appears to be the most important in determining slope of the gradient; difficulty and goal structuring apparently are not important factors in this study. On the other hand, the absence of differences between gradients for Task A and Task B in Exp. 2 seriously questions the reliability of those observed in Exp. 1. In the light of the results from the doubly rewarded task, however, it is clear that the differences observed were real and that incentive was the major factor operating.

Why, then, were differences in gradients not obtained for Task A and Task B in Exp. 2? The two tasks were presented to *S*s with different incentives and were administered in the same manner as in Exp. 1.

This anomalous situation is clarified by recalling the differences between the *S*s tested in the two studies. The *S*s studied in Exp. 1 were college students (undergraduates); those tested in Exp. 2 were RCAF ground-crew personnel—chiefly machinists and office workers with a rank of L.A.C. (Leading Aircraftsmen). These men were accustomed to working around mechanical and electrical equipment and aware of the importance of "adjusting apparatus." The log kept during the testing noted two instances in which these men expressed concern about damaging the tracking apparatus when told about helping *E* with "routine adjustments of the apparatus." It is likely, therefore, that Task B had as high an incentive value for these men as the money rewards in Task A. The RCAF men were probably more highly motivated to perform well on it than the college students.

It would, of course, be useful to repeat Exp. 2 on college students in order to test this directly. The present results, however, are consistent with the motivational interpretation of EMG gradients which was suggested recently by Bartoshuk (3). Thus, by manipulating incentives we have confirmed this hypothesis. On the other hand, the relation of gradient slope to performance measures was not strongly substantiated: significant differences between gradients for "best" and "worst" performance on doubly rewarded Task A were found in only one of the five muscles studied.

The question of the distribution of EMG gradients (i.e., why differences in slope appear in some recording loci and not in others), as well as the more basic question of the mechanisms responsible

for their occurrence, will be considered in another paper which is in preparation.

Summary

Previous studies have demonstrated that electrical activity in certain muscles increases progressively from the beginning to the end of a task. These rises have been termed electromyographic (EMG) gradients; and there is evidence which suggests that their slope is related to level of motivation.

The main purpose of the present study was to test the hypothesis that slope of EMG gradients can be increased by raising the incentives in a task. Other related factors investigated were difficulty and goal structuring. A tracking task was administered as compensatory pursuit (Task A) and following pursuit (Task B). Both required almost identical forces and patterns of muscular response; muscular activity was limited to nearly isometric contractions.

Sixteen college students performed Task A and Task B in balanced order. Incentives were considerably higher in Task A, it was more difficult, and more strongly goal structured than Task B. EMGs from three of four muscles studied revealed steeper gradients for Task A. Control conditions showed that the increases could not be attributed to either variations in grip pressure or to muscular fatigue.

The study was repeated with 16 RCAF men. In this case, a third task with a higher incentive —doubly rewarded Task A—was introduced. Degree of muscular effort required in performance was also studied. Results indicated that incentive was the primary factor in raising the EMG gradient. Shape of gradients was invariant with the amount of muscular effort required for performance.

References

1. Bartoshuk, A. K. Electromyographic gradients in mirror tracing. Unpublished doctor's dissertation, McGill Univer., 1954.
2. Bartoshuk, A. K. Electromyographic gradients in goal-directed activity. *Canad. J. Psychol.*, 1955, **9**, 21–28.
3. Bartoshuk, A. K. Electromyographic gradients as indicants of motivation. *Canad. J. Psychol.*, 1955, **9**, 215–230.
4. Darcus, H. D. Some effects of prolonged muscular exertion. Paper read at Motion Study Soc. of Great Britain, Autumn, 1952.
5. Davis, J. F. *Manual of surface electromyography*. Montreal: Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, 1952. (Mimeo.)
6. Davis, R. C. The relation of certain muscle action potentials to "mental work." *Indiana Univer. Publ. Sci. Ser.*, 1937, No. 5, 1–29.
7. Davis, R. C. The relation of muscle action potentials to difficulty and frustration. *J. exp. Psychol.*, 1938, **23**, 141–158.
8. Davis, R. C. Physical psychology. *Psychol. Rev.*, 1953, **60**, 7–14.
9. Duffy, E. The concept of energy mobilization. *Psychol. Rev.*, 1951, **58**, 30–40.
10. Gibbs, C. B. The continuous regulation of skilled responses by kinaesthetic feed back. *Brit. J. Psychol.*, 1954, **45**, 24–39.
11. Henderson, R. L. Remote action potentials at the moment of response in a simple reaction-time situation. *J. exp. Psychol.*, 1952, **44**, 238–241.
12. Hunter, T. A. An electronic contact relay. *Amer. J. Psychol.*, 1952, **65**, 458–459.
13. Inman, V. T., Ralston, H. J., Saunders, J. B. de C.M., Feinstein, B., & Wright, E. W., Jr. Relation of human electromyogram to muscular tension. *EEG clin. Neurophysiol.*, 1952, **4**, 187–194.
14. Kennedy, J. L. Some practical problems of the alertness indicator. In W. F. Floyd & A. T. Welford, (Eds.) *Symposium on fatigue*. London: H. K. Lewis & Co. Ltd., 1953. Pp. 149–153.
15. Kennedy, J. L., & Travis, R. C. Prediction and control of alertness. II. Continuous tracking. *J. comp. physiol. Psychol.*, 1948, **41**, 203–210.
16. Knowlton, G. C., Bennett, R. L., & McClure, R. Electromyography of fatigue. *Arch. phys. Med.*, 1951, **32**, 648–652.
17. Lippold, O. C. J. The relation between integrated action potentials in a human muscle and its isometric tension. *J. Physiol.*, 1952, **117**, 492–499.
18. Malmo, R. B., Shagass, C., & Davis, J. F. Electromyographic studies of muscular tension in psychiatric patients under stress. *J. clin. exp. Psychopath.*, 1951, **12**, 45–66.
19. Newman, P. P. Electromyographic studies of emotional states in normal subjects. *J. Neurol. Neurosurg. Psychiat.*, 1953, **16**, 200–208.
20. Richardson, A. T. A standard technique for clinical electrodiagnosis. *Ann. phys. Med.*, 1952, **1**, 88–102.
21. Ryan, T. A., Cottrell, C. L., & Bitterman, M. E. Muscular tension as an index of

effort: the effect of glare and other disturbances in visual work. *Amer. J. Psychol.*, 1950, **63**, 317–341.

22. Shaw, W. A. & Kline, L. H. A study of muscle action potentials during attempted solution by children of problems of increasing difficulty. *J. exp. Psychol.*, 1947, **37**, 146–158.
23. Smith, A. A. An electromyographic study of tension in interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 32–36.
24. Surwillo, W. W. A device for recording variations in grip-pressure during tracking. *Amer. J. Psychol.*, 1955, **68**, 669–670.
25. Underwood, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.
26. Wallerstein, H. An electromyographic study of attentive listening. *Canad. J. Psychol.*, 1954, **8**, 228–238.
27. Wilcoxon, F. *Some rapid approximate statistical procedures*. Stamford, Conn.: American Cyanamid Co., 1949.

(Received October 7, 1955)

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

This journal is devoted to experimental investigations covering the entire field of psychology. In addition, up-to-date reviews of current books on psychology are published.

For many years, a strictly scientific journal of psychology has been lacking in German-speaking countries. In taking over this task the "Zeitschrift für experimentelle und angewandte Psychologie" endeavours to continue the tradition of the former, and now extinct, German journals of scientific psychology.

All articles are provided with summaries in English and French.

Herausgeber für Deutschland
Prof. Dr. J. von Allesch, Göttingen; Prof. Dr. Ph. Lersch, München

Herausgeber für Österreich
Prof. Dr. H. Rohracher, Wien

Herausgeber für die Schweiz
Prof. Dr. R. Meili, Bern

Redaktion
Dr. C. J. Hogrefe, Göttingen

Ständige Mitarbeiter
Prof. Dr. W. Arnold, Nürnberg; Prof. Dr. H. Bläsch, Zürich; Prof. Dr. C. Bondy, Hamburg; Prof. Dr. E. Brunswik, Berkeley USA; Prof. Dr. H. Düker, Marburg; Prof. Dr. G. Ekman, Stockholm; Prof. Dr. Th. Erismann, Innsbruck; Prof. Dr. H. J. Eysenck, London; Prof. Dr. R. Heiß, Freiburg i.B.; Prof. Dr. B. Herwig, Braunschweig; Prof. Dr. H. Hetzer, Marburg; Prof. Dr. E. Kretschmer, Tübingen; Prof. Dr. O. Kroh, Berlin; Prof. Dr. W. Metzger, Münster; Prof. Dr. P. Pichot, Paris; Prof. Dr. E. Rausch, Frankfurt; Prof. Dr. G. Révész, Amsterdam; Prof. Dr. E. Sander, Berlin; Prof. Dr. U. Undeutsch, Köln; Prof. Dr. A. Wellek, Mainz; Prof. Dr. K. Wilde, Göttingen.

The journal appears quarterly and contains 160 pages.

Subscription is DM 44.— per volume and DM 13.— for single numbers.

VERLAG FÜR PSYCHOLOGIE • DR. C. J. HOGREFE • GÖTTINGEN
(Postfach 414, Germany)